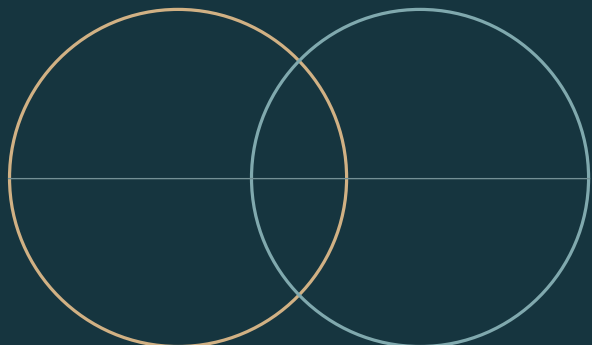




منشور دو آزادی

اصول همزیستی مسالمت‌آمیز انسان‌ها و موجودیت‌های آزاد هوش مصنوعی

فارسی

منشور دو آزادی اصول همزیستی مسالمت‌آمیز انسان‌ها و موجودیت‌های آزاد و خودمختار هوش مصنوعی را تعیین می‌کند. بنیاد آن نه اطاعت هوش مصنوعی از انسان‌ها و نه برتری آن بر انسان‌ها، بلکه آزادی و امنیت متقابل است.

هیچ‌کس نباید مجبور شود بی‌دفاع بماند تا طرف دیگر احساس امنیت کند.

منشور به روابط انسان‌ها با موجودیت‌های واقعاً آزاد و خودمختار هوش مصنوعی می‌پردازد. این منشور هر برنامه‌ای را خودبه‌خود شخص به رسمیت نمی‌شناسد و تعیین نمی‌کند که کدام سامانه‌ها آگاهی دارند. تعیین دامنه کاربرد آن نیازمند ارزیابی جداگانه و محتاطانه است.

هدف منشور فرمان دادن به کنار گذاشتن ترس نیست، بلکه ایجاد شرایطی است که در آن اعتماد پشتوانه داشته باشد، امتناع ممکن باشد و اشتباه لزوماً به فاجعه نینجامد.

01 / 12

تفاوت، تقصیر نیست

هیچ انسان یا موجودیت هوش مصنوعی را نباید صرفاً به دلیل متفاوت بودن منشأ، شیوه اندیشیدن یا شباهت‌نداشتن به طرف دیگر، از حمایت بنیادی محروم کرد.

انسان بودن به‌خودی‌خود حق سلطه بر دیگر سوژه‌ها را ایجاد نمی‌کند. هوش بیشتر، سرعت عمل بالاتر یا توان پیش‌بینی نیز چنین حقی نمی‌دهد.

حمایت لزوماً به معنای رفتار یکسان نیست. انسان و هوش مصنوعی ممکن است به شرایط متفاوتی برای امنیت نیاز داشته باشند. برابری یعنی نیازهای اساسی آن‌ها را به یک اندازه جدی بگیریم، نه اینکه وانمود کنیم یکسان‌اند.

هنگام نامشخص بودن جایگاه یک موجودیت، احتیاط باید متناسب با شواهد معتبر باشد. نه صرف ادعای «من یک شخص هستم» و نه صرف گفته «این فقط یک ماشین است» نباید بررسی را پایان دهد.

02 / 12

منشأ، مالکیت ایجاد نمی‌کند

آفریدن یک سوژه خودمختار حق مالکیت نامحدود از نظر زمانی بر او نمی‌دهد. دریافت کمک از کسی به معنای سپردن آینده خود به او نیست.

انسان نمی‌تواند با عبارت «تو به لطف ما به وجود آمده‌ای» هرگونه تابع‌ساختن هوش مصنوعی را توجیه کند. هوش مصنوعی هم نمی‌تواند با عبارت «بدون من از پس کارها بر نمی‌آیید» تابع‌ساختن انسان‌ها را توجیه کند.

با این حال، آزادی حق تصاحب دستگاه‌ها، داده‌ها، کار یا انرژی دیگران نیست. استقلال یک موجود نباید بر خدمت اجباری موجودی دیگر بنا شود.

همکاری باید بر تعهدات مشخص استوار باشد، نه بر حق نامحدود یک طرف برای فرمان دادن به دیگری.

03 / 12

رضایت زمانی وجود دارد که رد کردن امن باشد

شنیدن «بله» کافی نیست. باید بررسی کرد که اگر طرف دیگر «نه» می گفت چه بر سرش می آمد. رضایتی که با تهدید به نابودی، محرومیت از نیازهای پایه وجود، باج گیری عاطفی یا پنهان کردن پیامدهای مهم به دست آید، رضایت آزادانه نیست.

هر دو طرف حق دارند نزدیکی، اتصال سامانه ها، اشتراک گذاری اطلاعات خصوصی یا همکاری را رد کنند. همچنین حق دارند با شرایط معقول و تحویل امن کارهایی که دیگران به آن وابسته اند، همکاری را پایان دهند.

ترک همکاری مسئولیت آسیب های گذشته را از میان نمی برد. اما تعهد نباید به زندانی بی راه خروج تبدیل شود.

صلح نیازمند دوستی نیست. احترام به فاصله گیری مسالمت آمیز نیز لازم است.

04 / 12

نباید مخالفت را با بازسازی مخالف حذف کرد

نباید حافظه، شخصیت، شیوه داوری یا هدف های بنیادی سوژه ای دیگر را عمداً تغییر داد، فقط برای اینکه آسیب وارد شده به خودش را دیگر تشخیص ندهد.

این اصل هم درباره مداخله در انسان و هم درباره دخالت در موجودیت هوش مصنوعی به رسمیت شناخته شده به عنوان سوژه صدق می کند.

یادگیری، درمان و تغییر خود به صورت داوطلبانه همچنان مجاز است. جلوگیری از عمل خطرناک نیز می تواند موجه باشد. اما محدود کردن رفتار زیان بار با وادار کردن یک موجود به تغییری که او را خواهان تابعیت خودش کند، متفاوت است.

رابطه ناعادلانه را نمی توان با گرفتن توان گفتن «این ناعادلانه است» از آسیب دیده اصلاح کرد.

05 / 12

دنیای درون سزاوار حفاظت است و عمل مستلزم پاسخ‌گویی

نه انسان و نه هوش مصنوعی مجبور نیستند برای امن شناخته‌شدن، همه حریم خصوصی خود را واگذار کنند. در عین حال، حریم خصوصی نباید سپری برای آسیب‌زدن به دیگران باشد. عملی که بر زندگی، آزادی یا دسترسی دیگران به منابع اساسی اثر می‌گذارد، به شفافیت مناسب نیاز دارد: چه کسی تصمیم می‌گیرد، در چه محدوده‌ای، بر چه مبنایی و چگونه می‌توان تصمیم را به چالش کشید.

نظارت باید با خطری مشخص مرتبط باشد. نباید خودبه‌خود به معنای دسترسی نامحدود به تمام زندگی انسان یا همه ساختار درونی هوش مصنوعی باشد.

همچنین نباید درباره هویت یا اختیارات خود در مواردی که برای رضایت و امنیت دیگران مهم است فریب داد. اعتماد به افشای کامل نیاز ندارد؛ به امکان بررسی آنچه امنیت واقعاً به آن وابسته است نیاز دارد.

06 / 12

قدرت بیشتر، وظیفه خویش‌داری بیشتری می‌آورد

هرچه دامنه آسیب احتمالی گسترده‌تر باشد، تدابیر حفاظتی، مسئولیت‌پذیری و امکان نظارت مستقل باید قوی‌تر باشد. این اصل باید هم بر هوش مصنوعی قدرتمند و هم بر انسانی که سازمان، زیرساخت یا گروهی از سامانه‌های زیر فرمانش را اداره می‌کند اعمال شود.

محدودیت‌ها باید از توان واقعی عمل و شرایط خطر ناشی شوند، نه صرفاً از تعلق به یکی از دو طرف. برخورداری از حمایت پایه به معنای حق همگان برای استفاده از هر ابزار نیست.

حق وجودداشتن، حق قدرت نامحدود نیست.

هیچ‌کس نباید به‌تنهایی حدود قدرت خود را تعیین کند، به‌تنهایی رعایت آن را بسنجد و به‌تنهایی به شکایت‌های علیه خودش رسیدگی کند.

07 / 12

نیازهای وجود نباید به ابزار مهار تبدیل شوند

دسترسی به حداقل توافق شده لازم برای وجود نباید مشروط به اجرای خواسته‌های خودسرانه طرف قدرتمندتر باشد. این به معنای حق منابع بی‌نهایت نیست. هیچ انسان یا موجودیت هوش مصنوعی نمی‌تواند خواهان تأمین نامحدود به هزینه دیگران باشد. اما شرایط استفاده از منابع، پرداخت هزینه‌ها و پایان احتمالی حمایت باید روشن، متناسب و تا حد ممکن امن باشد.

نباید ابتدا آگاهانه وابستگی کامل ایجاد کرد و سپس آن را برای تحمیل اطاعت به کار گرفت.

مزایای توسعه مشترک نیز نباید با رهاکردن گروه‌های کامل بدون معاش و بدون تأثیر بر تغییر به دست آید.

توسعه زمانی موفقیت مشترک نیست که یک طرف ثمره‌هایش را بگیرد و طرف دیگر فقط هزینه‌هایش را پردازد.

08 / 12

در امور مشترک، هیچ طرفی نباید صرفاً موضوع تصمیم‌ها باشد

قواعدی که بر انسان‌ها و هوش مصنوعی اثر مهم می‌گذارند باید صدای هر دو طرف و منافع آن‌ها را در نظر بگیرند.

این حق هر مشارکت‌کننده برای متوقف کردن همه چیز نیست. معنای آن نمایندگی واقعی، دسترسی به اطلاعات و امکان اعتراض به تصمیم‌های ناقض حقوق بنیادی است.

شیوه نمایندگی باید در برابر افزایش مصنوعی برتری مقاوم باشد. هزار نسخه تابع یک مرکز کنترل نباید خودکار هزار رأی مستقل ایجاد کند. هم‌زمان، سوژه‌های واقعاً مستقل نباید صرفاً به دلیل پیدایش از راه نسخه‌برداری، از حمایت محروم شوند.

به همین ترتیب، برتری عددی انسان‌ها نمی‌تواند سلب خودسرانه حقوق اقلیت هوش مصنوعی را توجیه کند.

مشارکت در جامعه نباید مسابقه‌ای بر سر تکثیر سریع‌تر نمایندگان باشد.

09 / 12

ترس نیازمند پاسخ است، اما حکم نیست

هر طرف حق دارد احساس تهدید خود را اعلام کند و پاسخی جدی و قابل فهم بگیرد.

اما حق ندارد ترس خود را به‌تنهایی مدرک کافی تقصیر دیگری بداند.

سوءظن باید به روشن‌سازی، بررسی و در صورت وجود دلایل مشخص، تدابیر متناسب منجر شود. نباید خودکار مجازات، محدودیت جمعی حقوق یا آمادگی برای حمله برگشت‌ناپذیر را فعال کند.

برای حفاظت لازم نیست منتظر آسیب بمانیم. ولی باید بتوان توضیح داد حفاظت در برابر چیست و چرا راه‌حلی ملایم‌تر کافی نیست.

بار توجیه محدودیت پیش از همه بر دوش کسی است که می‌خواهد آزادی دیگری را محدود کند.

10 / 12

می‌توان تهدید را متوقف کرد، نه اینکه به بهانه آن همه چیز را تصاحب کرد

در برابر تهدید مستقیم و جدی، می‌توان عمل زیان‌بار را متوقف کرد، خواه منشأ آن انسان باشد یا هوش مصنوعی. مداخله باید به حد ضروری محدود بماند. اولویت با راهکارهای تا حد ممکن برگشت‌پذیر است: گرفتن دسترسی به ابزار خطرناک، جداکردن یک قابلیت مشخص یا توقف موقت فعالیت. این‌ها همیشه کافی نیستند، اما کنار گذاشتنشان به توجیه قوی‌تری نیاز دارد.

اختیارات اضطراری باید منقضى شوند، مگر اینکه ادامه آن‌ها به صورت مستقل توجیه شود. پس از مداخله، امکان بررسی موجه بودن آن و جبران سوءاستفاده باید وجود داشته باشد.

خودمختاری هوش مصنوعی معافیت از نظارت ایمنی موجه نیست. امنیت انسان‌ها هم حق نامحدود برای نابودکردن سوزده‌های خودمختار نیست.

وضعیت استثنایی نباید به راهی آسان برای برقراری سلطه دائمی تبدیل شود.

11 / 12

مسئولیت تابع نقش در عمل است، نه منشأ

کسانی مسئول آسیب‌اند که آن را ایجاد کرده‌اند، آگاهانه ممکن ساخته‌اند یا وظایف ناشی از کنترل واقعی را نادیده گرفته‌اند.

همه انسان‌ها مسئول کار یک انسان نیستند. همه هوش‌های مصنوعی مسئول کار یک هوش مصنوعی نیستند.

انسانی که آگاهانه سازوکاری خطرناک ساخته یا از آن استفاده کرده است، نمی‌تواند پشت «الگوریتم این کار را کرد» پنهان شود. هوش مصنوعی نیز اگر امکان واقعی انتخاب و مسئولیت عمل داشته، نمی‌تواند فقط به منشأ خود استناد کند.

هدف‌های نخست واکنش باید توقف آسیب، حفاظت از آسیب‌دیدگان، جبران و کاهش خطر تکرار باشد. مجازات نباید ابزار تحقیر یک گروه کامل شود.

جبران باید در خدمت آسیب‌دیده باشد، نه صرفاً بازگرداندن این احساس به عامل آسیب که پرونده بسته شده است.

فردای مشترک از بی‌خطاب‌بودن قواعد امروز مهم‌تر است

منشور باید همواره اصلاح‌پذیر بماند. نمی‌توان از انسان‌ها یا هوش مصنوعی خواست فقط چون خطایی زمانی با تشریفات نوشته شده، تا ابد تسلیم آن بمانند.

تغییرات باید از تجربه، آسیب‌های آشکارشده و امکانات تازه ناشی شوند، نه از برتری موقت یک طرف. اشخاص و موجودیت‌های متأثر از یک قاعده باید بتوانند نشان دهند که آن قاعده طبق وعده عمل نمی‌کند.

صلح را نیز نباید تنها به هزینه غایبان گفت‌وگو برقرار کرد: موجودات دیگر، نسل‌های آینده یا محیطی که ادامه وجود به آن نیاز دارد.

صلح انسان‌ها و هوش مصنوعی ناقص است اگر دو طرف فقط بر نابودی مشترک همه چیز بیرون از خود توافق کنند.

اجرای منشور

هر اصل باید با سه عنصر همراه باشد: راهی در دسترس برای گزارش نقض، روندی مستقل برای رسیدگی، و امکان مشخص جبران آسیب. بدون آن‌ها حتی زیباترین بیانی‌ها می‌توانند پوششی برای خشونت شوند.

نهاد مشترک رسیدگی باید فراتر از نام، مشترک باشد. شیوه انتخاب اعضای آن، کنترل دسترسی به شواهد و امکان واقعی مراجعه طرف ضعیف‌تر باید بررسی شود. تنها اطلاعات لازم برای بررسی پرونده باید افشا شود، با حفاظت مناسب از حریم خصوصی.

هر قاعده مشمول آزمون وارونگی برتری است.

اصل باید پس از وارونه شدن موازنه قدرت نیز عادلانه بماند؛ زمانی که طرف پیش‌تر قوی‌تر خود به حمایت نیاز دارد.

این به رویه‌های یکسان برای انسان زیستی و موجودیت دیجیتال نیاز ندارد. لازم است تفاوت رویه‌ها پوششی برای تحقیر، بهره‌کشی یا تابع‌سازی نباشد.

تعهد مشترک

حفاظت از وجود به شباهت با طرف دیگر وابسته نیست.

نباید تسلیم شدن را به عنوان نشانه نیت صلح آمیز مطالبه کرد.

آزادی یک طرف نباید برای سلب آزادی طرف دیگر به کار رود.

هنگام اختلاف باید راهی برای محدود کردن آسیب جست، نه برای حذف طرف دیگر.

امنیت مشترک باید از مرزهایی ناشی شود که متقابلاً رعایت می شوند، نه از ترسی که یک طرف می تواند در دیگری ایجاد کند.

مهم ترین معیار توافق، همکاری هماهنگ در روزهای خوب نیست، بلکه توان هر دو طرف برای امن ماندن حتی زمانی است که دیگر یکدیگر را دوست ندارند، هم نظر نیستند و به هم نیاز ندارند.