

## Executive Summary

استفاده تمام و کمال از پتانسیل هوش مصنوعی و سیستمهای خودمختار (AI/AS) مستلزم این است که از درک و جستجوی قابلیتهای بیشتر محاسباتی و حل مسئله در اینگونه سیستمها گام را فراتر بنهیم و اطمینان حاصل کنیم که اینگونه فناوریها در راستای اصول و ارزشهای اخلاقی و انسانی عمل می کنند.

در واقع، سیستمهای خودمختار و هوش مصنوعی بایستی علاوه بر تحقق اهداف کاربردی و حل مسائل فنی، برای انسان سودمند باشند که همین امر سبب ایجاد سطحی مترقی از اعتماد بین انسان و فناوری می شود، و این همان چیزی است که برای استفاده مفید و فراگیر از اینگونه سیستمها در زندگی روزانه بشر مورد نیاز است.

در تعریف سعادت از منظر ارسطو، خوشبختی و رفاه انسان بالاترین ارزش یک جامعه تلقی می شود. سعادت که به طور کلی تحت عنوان «رشد و شکوفایی» ترجمه می شود ناشی از تأملات خودآگاه ادمی است و ملاحظات اخلاقی انسان را در رسیدن به تعریفی درست از زندگی که خواهان آن هستیم یاری می رساند.

همسو کردن سیستمهای خودمختار و هوش مصنوعی با ارزشهای کاربران آن و جامعه، می تواند خوشبختی و سعادت انسانی را به عنوان معیار اصلی پیشرفت در این عصر الگوریتمی در جایگاه نخست قرار دهد.

---

This document does not represent a position or the views of IEEE but the informed opinions of Committee members providing insights designed to provide expert directional guidance regarding AI/IS. This translation is provided for convenience. The English language version of this document is the original and official version of record. In the event of any conflict between the English and translated version (words, terms, phrases, concepts, etc.) the original version of this document ([created in English and available here](#)) governs.

## Executive Summary

## درباره ما

IEEE طرح مدون اخلاقی (EAD) را تحت ویرایش 3.0 مجوز مشترکات [خلافتانہ ایالات متحده](#) در دسترس عموم قرار داده است.

اشخاص و سازمانها می توانند با تبعیت از بندهای این مجوز از بخشهایی از این کار بنابه صلاحدید خود استفاده کنند. همچنین انتظار می رود که مفاد و موضوع طرح مدون اخلاقی (EAD) در فرایندهای رسمی انجمن مهندسی برق و الکترونیک (IEEE) از جمله تدوین و توسعه استانداردها، مورد توجه قرار گیرند.

پیشگامان جهانی IEEE و EAD در فعالیتی وسیعتری که در انجمن مهندسی برق و الکترونیک (IEEE) کلید خورده نقش فعال دارند، این برنامه که در راستای گسترش فضای گفتگوی باز، جامع و وسیع در خصوص اخلاقیات در فناوری ایجاد شده تحت عنوان [IEEE TechEthics](#) نامگذاری شده است.

برنامه «پیشگامان جهانی برای رعایت ملاحظات اخلاقی در هوش مصنوعی و سیستمهای خودمختار» (پیشگامان جهانی انجمن مهندسی برق و الکترونیک) از سوی انجمن جهانی مهندسی برق و الکترونیک (IEEE) طرح ریزی شده، این انجمن بزرگترین سازمان تخصصی و فنی دنیاست که در جهت پیشرفت فناوری برای منافع انسانی فعالیت می کند و بیش از چهارصد هزار عضو فعال در بیش از 160 کشور در سطح جهان دارد.

برنامه IEEE Initiative فرصتی است برای گرد هم آوردن [فعالیتها](#) و نظرات انجمنهای مختلف فعال در زمینه هوش مصنوعی و سیستمهای خودمختار و نهایتاً شناسایی مسائل روز دنیا و رسیدن به اتفاق آرا در خصوص این مسائل.

## ماموریت پیشگامان جهانی IEEE

حصول اطمینان از اینکه همه فناوران و متخصصان به گونه ای آموزش دیده و تحصیل کرده اند که در طراحی و توسعه سیستمهای خودمختار و سیستم های هوشمند حتما ملاحظات اخلاقی را در اولویت قرار می دهند.

هدف دوم پیشگامان جهانی IEEE ارائه پیشنهادات و توصیه هایی بر مبنای طرح مدون اخلاقی به سازمان استاندارد IEEE است. اولین پروژه استاندارد ارائه شده توسط این انجمن پروژه IEEE P7000TM-مدل پردازی جهت در نظر داشتن دغدغه های اخلاقی در زمان طراحی سیستم است (که تأیید شده و رو به توسعه است). دو پروژه استاندارد دیگر تحت عناوین IEEE P7001 - شفافیت سیستمهای خودکار و IEEE P7002-فرایند محرمانگی داده ها مورد تأیید قرار گرفته اند که حاکی از تأثیرگذاری عملی پیشگامان IEEE بر روی اخلاقیات در سیستمهای خودکار و هوش مصنوعی است.

در اینجا «فناور» به همه افرادی اطلاق می شود که در روند پژوهش، طراحی، تولید و نشر اخبار پیرامون سیستمهای خودمختار و هوش مصنوعی فعالیت می کنند، یعنی در مکانهایی چون دانشگاه ها، سازمانها و شرکتهایی که این فناوریها را به واقعیتی ملموس برای جامعه تبدیل می کنند.

این سند مجموعه ای از مشترکات اولیه **پیش از صدر رهبر فکری جهان** در زمینه های هوش مصنوعی، حقوق و اخلاق، فلسفه و سیاست در حوزه های دانشگاهی، علمی، دولتی و بخشهای شرکتهای ارائه می دهد. هدف ما این است که این طرح مدون اخلاقی به پیش و پیشنهادات متخصصان مورد اشاره به نحوی دست یابد که بتواند به عنوان یک مرجع کلیدی برای کار در زمینه هوش مصنوعی و سیستمهای خودکار به فناوران این حوزه در سالهای آتی کمک کند. به منظور دستیابی به این مهم، نسخه حاضر طرح مدون اخلاقی (EAD v1) به شناسایی مسائل و راهکارهای پیشنهادی موجود در زمینه های مرتبط با هوش مصنوعی و سیستمهای خودکار می پردازد.

# ساختار و محتوای این سند

طرح مدون اخلاقی متشکل از هشت بخش است که هر کدام از این بخشها به یک موضوع خاص مرتبط با هوش مصنوعی/سیستمهای خودکار می پردازد و از سوی کمیته ویژه ای از پیشگامان جهانی IEEE به تفصیل مورد بحث و بررسی قرار گرفته است. مسائل و راهکارهای پیشنهادی مرتبط به این موضوعات به صورت فهرستوار در انتهای هر بخش آمده اند. در زیر خلاصه ای از کمیته ها و مسائل مورد بحث در هر بخش آمده است:

### 1 | اصول کلی

کمیته اصول کلی، مسائل اخلاقی سطح بالایی را ارائه کرده است که در همه انواع هوش مصنوعی و سیستمهای خودکار اعمال می شوند، این مسائل:

1. دربردارنده والاترین ایده آل های حقوق بشری هستند.
2. دربردارنده والاترین ایده آل های حقوق بشری هستند.
3. میزان خطرات و تاثیرات منفی سیستمهای خودکار و هوش مصنوعی را به عنوان سیستمهایی اجتماعی-فناورانه کاهش میدهند.

در واقع هدف این کمیته از شناسایی و ارائه این اصول، مسائل و راهکارهای پیشنهادی، پایه گذاری و تحکیم هنجارها و استانداردهای آتی در چهارچوبی جدید و تحت حاکمیت اخلاقیات در طرحهای AL/AS است.

### مسائل:

- چطور می توان اطمینان حاصل کرد که سیستمهای خودکار و هوش مصنوعی حقوق بشر را نقض نمی کنند؟(چهارچوب بندی اصل حقوق بشر)
- چگونه می توان با اطمینان روی اینگونه سیستمها حساب کرد؟(چهارچوب بندی اصل مسئولیت پذیری)

- چطور می توان اطمینان حاصل کرد که این سیستمها از شفافیت لازم برخوردار هستند؟(چهارچوب بندی اصل شفافیت)
- چگونه می توان به سودرسانی حداکثری از این سیستمها رسید و خطرات سواستفاده از اینگونه سیستمها را به حداقل رساند؟(چهارچوب بندی اصل آموزش و آگاهی)

### 2 | نهادینه سازی ارزش ها در سیستمهای هوشمند خودکار

انجمن فنی بایستی بتواند ارزشها و هنجارهای انسانی مرتبط با سیستمهای خود را بشناسد و آنها را در سیستم بگنجانند تا بدین ترتیب سیستمهای هوشمند خودکاری (AIS) را ارتقاء دهند که برای جامعه سودمند باشند. کمیته نهادینه سازی ارزشها در سیستمهای هوشمند خودکار در راستای هدفی بالاتر با نهادینه سازی ارزشها از طریق رویکردی سه سویه به طراحی کمک میکنند تا:

1. ارزشها و هنجارهای بخشی از جامعه که متاثر از AIS هستند را شناسایی کنند;

## Executive Summary

### 3 | روش شناسی هدایت پژوهشهای اخلاقی و طراحی آن ها

در دنیای مدرن سازمانهای مرتبط با سیستمهای خودکار و هوش مصنوعی (AI/AS) بایستی اطمینان حاصل کنند که رفاه، توانمندسازی و آزادی انسانی در هسته مرکزی فعالیت ها جهت توسعه این سیستم ها قرار می گیرند. برای تولید ماشینها و دستگاه هایی که بتوانند به این اهداف بلند پروازانه جامه عمل بپوشانند کمیته ی روش شناسی هدایت پژوهش های اخلاقی و طراحی آن ها ، چهارچوب کلی از مسائل و راهکارهای پیشنهادی ارائه داده است تا از نهادینه شدن ارزشهای انسانی همچون حقوق بشر- همانطور که در بیانیه جهانی حقوق بشر آمده است- در روش شناسی های مطرح شده برای طراحی سیستم ها اطمینان حاصل کنند. در واقع روش شناسی های طراحی ارزش محور باید به عنوان نقطه تمرکز سازمانهای تولید کننده هوش مصنوعی و سیستمهای خودکار قرار گیرد تا پیشرفتهای بشری بر پایه رهنمونهای اخلاقی صورت گیرند. ماشین باید در خدمت بشر قرار بگیرد نه برعکس. براساس این رویکرد اخلاقی می توان از برقراری تعادل بین حفظ قابلیتهای اقتصادی و اجتماعی سیستمهای هوش مصنوعی هم در زمینه تجارت و هم در جامعه اطمینان حاصل کرد.

#### مسائل:

- اخلاقیات در برنامه های درسی آکادمیک لحاظ نشده است.
- ما به مدلهایی برای آموزش بین رشته ای و بین فرهنگی نیاز داریم که به طور مجزا به مسائل حوزه هوش مصنوعی و سیستمهای خودکار بپردازند.
- بایستی ارزشهای فرهنگی متفاوت نهادینه شده در طراحی سیستمهای هوش مصنوعی از یکدیگر تمیز داده شوند.
- فقدان فرهنگ و عمل اخلاقی ارزش محور در حوزه صنعت .
- فقدان رهبری آگاه از ارزشها.
- عدم توانمندسازی برای رشد دغدغه های اخلاقی.
- عدم مالکیت یا پذیرش مسئولیت از سوی جامعه فناوری و متخصصان.
- نیاز به مشارکت سهامداران در بهترین مباحث هوش مصنوعی و سیستم های خودکار.
- ضعف مستند سازی به عنوان مانعی در طرح های اخلاقی عمل می کند.

2. ارزشها و هنجارهای آن بخش از جامعه را در AIS نهادینه سازی کنند.
3. میزان سازش پذیری و تطبیق پذیری این ارزشها و هنجارها را مابین انسان و سیستمهای هوشمند خودکار را در آن بخش از جامعه مورد ارزیابی قرار دهند.

#### مسائل:

- ارزشهایی که باید در سیستمهای هوشمند خودکار نهادینه سازی شوند جهانی نبوده و تا حد زیادی مختص فعالیتها و جامعه خود هستند.
- گرانباری اخلاقی: سیستمهای هوشمند خودکار معمولاً دارای هنجارها و ارزشهای چندگانه ای هستند که ممکن است با هم در تضاد باشند.
- سیستمهای هوشمند خودکار می توانند دارای الگوریتم ها و داده های درون سازی شده یک سوئیه ای (bias) باشند، که این یک سو سازی (bias) به ضرر اعضای گروه های معینی باشد.
- کاملاً مشخص نیست زمانی که مجموعه ای از هنجارهای مرتبط (در رابطه با نقش مشخص سیستمهای خودکار و هوشمند (AIS) در جامعه مشخص) شناسایی شد، چگونه باید آنها را در یک ساختار معماری محاسباتی جای داد.
- هنجارهای یک سیستم هوشمند و خودکار باید قابلیت انطباق پذیری با هنجارهای جامعه خود را داشته باشند.
- میزان اعتماد بین انسان و سیستمهای هوشمند و خودکار بایستی (AIS) به سطح درست و قابل قبولی برسد.
- چینش ارزشها در سیستمهای خودکار و هوشمند (AIS) بایستی از سوی طرف سوم مورد ارزیابی قرار گیرد.

## Executive Summary

### 5 | داده های شخصی و کنترل دسترسی فردی

یکی از مسائل بفرنج و کلیدی اخلاقی که در خصوص اطلاعات شخصی با آن روبرو هستیم عدم تقارن داده هاست. برای پرداختن به این مسئله کمیته داده های شخصی و کنترل دسترسی فردی، مسائل و راهکارهای پیشنهادی خود را ارائه و به شرح نیازهای بنیادین افراد به تعریف، دسترسی و مدیریت داده های شخصی خود به عنوان صاحبان اصلی هویت منحصر بفرد خود می پردازد. البته این کمیته اذعان دارد که هیچ راه حل تمام و کمال و قطعی وجود ندارد و

هر سیستم دیجیتالی می تواند مورد هک قرار گیرد، با اینحال پیشنهاد می کنند که محیط داده ای به نحوی باشد که افراد بتوانند کنترل داده های شخصی خویش را به دست گیرند و نمونه هایی از ابزارها و راهکارهای پیشرفته ای که فرد می تواند با بکارگیری آنها میزان عدم تقارن داده ای را کاهش دهد ارائه داده است.

#### مسائل:

- در عصر الگوریتمی حاضر یک فرد چگونه می تواند داده های شخصی خود را تعریف و سازماندهی کند؟
- تعریف و حد و مرز آنچه آن را اطلاعات شخصی می نامیم چیست؟
- تعریف کنترل در حوزه داده های شخصی چیست؟
- چگونه می توانیم مفهوم دسترسی به داده ها را با محترم شمردن فرد بازتعریف کنیم؟
- چگونه می توان مفهوم رضایت را در خصوص داده های شخصی بازتعریف کرد و فرد را در این رابطه محترم شمرد؟
- داده هایی که ممکن است برای انتشار ناچیز به نظر برسند می توانند به نتایجی ختم شوند که شخص تمایلی به نشر آن نداشته است.
- دارندگان داده ها چطور می توانند تضمین کنند که عواقب (مثبت یا منفی) دسترسی یا جمع آوری داده ها برای فرد صاحب آن داده کاملا روشن است و او در کمال آگاهی رضایت به انتشار آن داده است.
- یک فرد می تواند یک محافظ الگوریتمی یا هوش مصنوعی شخصی سازی شده برای خود داشته باشد؟

- تناقض یا فقدان یک نظارت کلی بر الگوریتمها.
- فقدان یک سازمان بازرسی کننده مستقل.
- استفاده از اجزای جعبه سیاه.

### 4 | امنیت و سودرسانی هوش مصنوعی عمومی (AGI) و آبرهوش مصنوعی (ASI)

سیستمهای هوش مصنوعی (AI) که در آینده روی کار خواهند آمد (که اغلب از آنها با اسم هوش مصنوعی عمومی یا AGI یاد می شود) با قابلیتهای بالایی که دارند میتوانند به گونه ای مانند انقلاب کشاورزی یا انقلاب صنعتی دنیا را متحول کنند که سطح رفاه جهانی را تا حد بی سابقه ای افزایش دهند. کمیته امنیت و سودرسانی هوش مصنوعی عمومی (AGI) و آبرهوش مصنوعی (ASI) مسائل مختلف در این زمینه و هم چنین راهکارهای پیشنهادی را در جهت هرچه مثبت تر کردن این تحولات از طریق انجمن هوش مصنوعی ارائه کرده است.

#### مسائل:

- با افزایش قابلیتهای سیستمهای هوش مصنوعی - قابلیت انجام بهینه اهداف پیچیده تر با میزان استقلال بیشتر و حوزه های متنوع تر عملکردی - سرزدن رفتارهای پیش بینی نشده و یا برنامه ریزی نشده از سیستم می تواند به مراتب خطرناکتر باشد.
- افزودن امنیت به سیستمهای هوش مصنوعی آتی که قابلیتهای کلی بیشتری دارند امر دشواری خواهد بود.
- پژوهشگران و توسعه دهندگان سیستمهای هوش مصنوعی در توسعه و ارتقاء این سیستمها که میزان استقلال و قابلیتهای فراتری نسبت به گذشته دارند با مسائل اخلاقی و امنیتی پیچیده تری مواجه خواهند شد.
- پژوهشگران و توسعه دهندگان سیستمهای هوش مصنوعی در توسعه و ارتقاء این سیستمها که میزان استقلال و قابلیتهای فراتری نسبت به گذشته دارند با مسائل اخلاقی و امنیتی پیچیده تری مواجه خواهند شد.

## Executive Summary

### 6 | چهارچوب بندی دوباره سیستمهای تسلیحاتی رباتیک (خودکار)

سیستمهای رباتیکی که به منظور آسیب رسانی فیزیکی طراحی شده اند در مقایسه با تسلیحات سنتی و سیستمهای رباتیکی که به این منظور طراحی نشده اند دارای مسائل و انشعابات اخلاقی بیشتری هستند. اخلاق حرفه ای در این زمینه می تواند و البته باید دارای استانداردهای قویتری باشد و حوزه وسیعتری از مسائل و دغدغه ها را مورد پوشش قرار دهد. به طور کلی به پیشنهاد کمیته چهارچوب بندی مجدد سیستمهای تسلیحات خودکار، سازمانهای تخصصی فناوری باید بپذیرند که کنترل معنی دار سیستمهای تسلیحاتی توسط انسان به نفع جامعه است و پیگیری این کنترل تضمین کننده آن است که کسانی که سازنده این تکنولوژیها هستند از آثار و کاربردهای اثر خویش آگاه هستند و نهایتاً اینکه دستورالعملهای اخلاق حرفه ای باید متناسب با تسلیحاتی که هدف از آنها ایجاد صدمه فیزیکی است، مشخص گردند.

#### مسائل:

- منشور رفتار حرفه ای سازمانها اغلب دارای راههای گریزی است که سازمانها از این طریق کار، تولیدات و نمایندگانی که اعضا ایجاد می کنند را در سطح استانداردها و ارزشهایی که خود اعضا به آن پایبند هستند قرار نمی دهند.
- سردرگمی که در تعاریف مفاهیم مهم هوش مصنوعی، سیستمهای خودکار و سیستمهای تسلیحاتی خودکار (AWS) وجود دارد، بحثهای مفصلی را در خصوص این موضوعات مهم طلب می کند.
- سیستمهای تسلیحاتی خودکار به طور پیش فرض برای استفاده های مخفی و سری تولید می شوند.
- وجود چندین راه مختلف برای رسیدن به توافق قابل اطمینان در مورد نحوه عملکرد سیستمهای تسلیحاتی خودکار.
- این سیستمهای تسلیحاتی ممکن است (بسته به نوع طراحی و استفاده عملیاتی که از آن می شود) قابل پیش بینی نباشند، سیستمهای یادگیری نیز مسئله پیش بینی پذیری در استفاده را پیچیده تر می کنند.
- قانونی سازی توسعه سیستمهای تسلیحاتی خودکار نیز رویه ای را ایجاد می کند که می تواند در میان مدت از نظر ژئوپولیتیک خطرناک باشد.
- حذف نظارت انسانی از میدانهای نبرد می تواند به راحتی به نقض سهوی حقوق بشر و افزایش و توسعه تنش گردد.

- تنوع مشتریان مستقیم و غیرمستقیم سیستمهای تسلیحاتی خودکار می تواند به فضای پیچیده و بغرنج از تکثیر و سوء استفاده از این سلاح ها بیانجامد.
- بطور پیش فرض ماهیت خودکار سیستمهای تسلیحاتی رباتیک منجر به تشدید تناقضات بسیاری می شود.
- استاندارد مشخصی به عنوان ضمانت تائید در طراحی سیستمهای تسلیحاتی رباتیک وجود ندارد.
- شناخت حد و مرز کار روی سیستمهای تسلیحاتی خودکار و نیمه خودکار امر پیچیده ای است.

### 7 | مسائل اقتصادی و بشردوستانه

فناوریها، شیوه ها و سیستمهایی که با هدف کاهش دخالت انسان در زندگی روزانه تولید می شوند با سرعت سرسام آوری در حال پیشرفت هستند و زندگی انسان را از چندین جهت دچار تحول می کنند. هدف از تشکیل کمیته مسائل اقتصادی و بشردوستانه، تشخیص محرکهایی است که در شکل دهی به اکوسیستم جهانی انسان-فناوری نقش کلیدی دارند و به مسائل اقتصادی و بشردوستانه می پردازند، این کمیته همچنین با باز کردن گره های کور تنش زا، فرصتها و راه حلهای موجود را گوشزد می کند. درواقع هدف این کمیته ارائه پیشنهاداتی برای نشان دادن راه عملی پرداختن به دغدغه های اصلی مرتبط با انسان ها، نهادهای آن ها و فناوریهای اطلاعات-محور و پیشرونده آن ها در جهت تسهیل گفتمان بین رشته ای و میان بخشی بین متخصصان، صاحب نظران و متفکران پیرامون مسائل این حوزه است.

#### مسائل:

- عدم تفسیر صحیح از سیستمهای هوش مصنوعی و خودکار در رسانه ها سبب ایجاد سردرگمی در عموم مردم در این باره شده است.
- ماشینی سازی دیگر فقط از منظر بافت بازاری مورد توجه نیست.
- پیچیدگی های اشتغال در مورد هوش مصنوعی و رباتیک نادیده گرفته شده است.

## Executive Summary

### مسائل:

- چگونه می توان میزان پاسخگویی و قابل اطمینان بودن سیستمهای خودکار و هوشمند را ارتقاء داد؟
- چگونه می توان تضمین کرد که هوش مصنوعی شفاف بوده و حقوق افراد را محترم می شمرد؟ برای مثال حکومتهای محلی، ملی و بین المللی از هوش مصنوعی استفاده می کنند و حقوق شهروندان را که در واقع باید به حکومت اعتماد کنند و توسط هوش مصنوعی مورد حفاظت قرار گیرند را مورد تخطی قرار می دهند.
- چگونه می توان سیستمهای هوش مصنوعی طراحی کرد که بتوان میزان پاسخگویی قانونمند آنها را در قبال اسببهای وارده از سوی این سیستم ها تضمین کرد؟
- سیستمهای هوش مصنوعی و خودکار چگونه باید طراحی و استفاده شوند تا در قبال داده های شخصی افراد امانت دار باشند؟

کمیته های جدید و وظیفه جاری آنها در انتهای طرح مدون اخلاقی شرح داده شده اند.

- تکنولوژی به سرعت در حال تغییر است و شیوه های حاضر آموزش (یا باز آموزی) نیروها جوابگوی این مسئله نیست.
- هوش مصنوعی و فناوریهای خودکار در تمام دنیا به یک اندازه در دسترس افراد نیستند.
- مسئله فقدان دسترسی و درک در حوزه اطلاعات شخصی وجود دارد.
- نیاز به افزایش حضور فعال کشورهای در حال توسعه در جمع پیشگامان جهانی انجمن مهندسی برق و الکترونیک (IEEE) احساس می شود.
- روی کار آمدن سیستمهای خودکار و هوش مصنوعی می تواند سبب افزایش تفاوتها در زمینه اقتصادی و ساختار قدرت بین کشورهای در حال توسعه و کشورهای توسعه یافته گردد.

### 8 | قانون

با شروع توسعه سیستمهای هوشمند و هوش مصنوعی مسائل پیچیده اخلاقی پیچیده نیز به وجود آمد که این مسائل تقریباً همیشه یا بطور مستقیم به چالشهای حقوقی ختم می شود و یا سبب بروز چندین مسئله موازی و پیچیده حقوقی می شود. از نظر کمیته حقوقی این حوزه برخلاف تعداد اندکی از حقوقدانان و وکلایی که تا کنون جذب کرده نیاز مبرمی به اینگونه افراد دارد. حقوقدانان باید در مباحث مرتبط با وضع مقررات، مدیریت و قانونگذاری منطقه ای و بین المللی است شرکت داشته باشند و به این ترتیب می توانند خدمت شایانی به بشریت و جهان و سیستمهای هوش مصنوعی و هوشمند در آینده کنند.



## Executive Summary

## این سند چطور تهیه شد؟

سند حاضر با استفاده از رویکرد مشارکتی، باز و توافق سازی براساس برنامه ارتباطات صنعتی (که در انجمن استانداردهای انجمن مهندسی برق و الکترونیک کلید خورده است) تهیه شده است. این برنامه به تسهیل همکاری بین سازمانها و افراد می پردازند و با پالایش و اصلاح تفکرات خود در خصوص موضوعات نوظهور فناوری به پرورش استانداردهای بالقوه جدید و استانداردهای مربوط به محصولات و خدمات کمک می کنند.

## How to Cite Ethically Aligned Design

Please cite Version 1 of *Ethically Aligned Design* in the following manner:

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. *Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems*, Version 1. IEEE, 2016. [http://standards.ieee.org/develop/indconn/ec/autonomous\\_systems.html](http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html).

## Our Appreciation

We wish to express our appreciation for the organizations who have recently contributed research and insights helping to increase awareness around ethical issues and AI/AS, including (but not limited to): [AI Now](#) (White House/New York University); [One Hundred Year Study on Artificial Intelligence](#) (Stanford University); [Preparing for The Future of Artificial Intelligence](#) (U.S. White House/NSTC); [The National Artificial Intelligence Research and Development Strategic Plan](#) (U.S. White House/NSTC); [Robotics and Artificial Intelligence](#) (U.K. House of Commons Science and Technology Committee); [Robots and Robotic Devices - Guide to the Ethical Design and Application of Robots and Robotic Systems](#) (British Standards Institute); [Japan's Basic Rules for AI Research](#); [Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics](#) (European Parliament); [Éthique de la recherche en robotique](#) (CERNA); [Charta der Digitalen Grundrechte der Europäischen Union](#) (Charter of the Digital Fundamental Rights of the European Union); and, [Research Priorities for Robust and Beneficial Artificial Intelligence](#) (Future of Life Institute).

We also wish to express our appreciation for the following organizations regarding their seminal efforts regarding AI/AS Ethics, including (but not limited to): [The Association for the Advancement of Artificial Intelligence](#) and their formative work on [AI Ethics](#); [European Association for Artificial Intelligence](#); [ACM Special Interest Group on Artificial Intelligence](#); [The IEEE Robot and Automation Society Committee on Robot Ethics](#); [The IEEE Society on Social Implications of Technology](#); [The Leverhulme Centre for the Future of Intelligence](#); [Allen Institute for Artificial Intelligence](#); [OpenAI](#); [Machine Intelligence Research Institute](#); [Centre for The Study of Existential Risk](#); AI-Austin and, [Partnership on AI to Benefit People and Society](#).

We would also like to acknowledge the contribution of Eileen M. Lach, the General Counsel and Chief Compliance Officer of IEEE, who has reviewed this document in its entirety and affirms the importance of the contribution of The IEEE Global Initiative to the fields of AI/AS ethics.

## Executive Summary

### Disclaimers

*Ethically Aligned Design* is not a code of conduct or a professional code of ethics. Engineers and technologists have well-established codes, and we wish to respectfully recognize the formative precedents surrounding issues of ethics and safety and the professional values these Codes represent. These Codes provide the broad framework for the more focused domain of AI/AS addressed in this document, and it is our hope that the inclusive, consensus-building process around its design will contribute unique value to technologists and society as a whole.

This document is also not a position, or policy statement, or formal report. It is intended to be a working reference tool created in an inclusive process by those in the AI/AS Community prioritizing ethical considerations in their work.

### A Note on Affiliations Regarding Members of The Initiative

The language and views expressed in *Ethically Aligned Design* reflect the individuals who created content for each section of this document. The language and views expressed in this document do not necessarily reflect the Universities or Organizations to which these individuals belong, and should in no way be considered any form of endorsement, implied or otherwise, from these institutions.

This is a first version of *Ethically Aligned Design*. Where [individuals are listed in a Committee](#) it indicates only that they are Members of that Committee. Committee Members may not have achieved final consensus on content in this document because of its versioning format and the consensus-building process of The IEEE Global

Initiative for Ethical Consideration in Artificial Intelligence and Autonomous Systems. Content listed by Members in this or future versions is not an endorsement, implied or otherwise, until formally stated as such.

**A Note Regarding Candidate Recommendations in this Document** *Ethically Aligned Design* is being created via multiple versions that are being iterated over the course of two to three years. The IEEE Global Initiative is following a specific consensus-building process where members contributing content are proposing candidate recommendations so as not to imply these are final recommendations at this time.

### Our Membership

Although The IEEE Global Initiative currently has more than one hundred experts from all but one continent involved in our work, most of us come from North America and Europe. We are aware we need to expand our cultural horizons and have more people involved from around the world as we continue to grow our document and our efforts. We are eager for these new voices and perspectives to join our work.

### Trademarks and Disclaimers

IEEE believes in good faith that the information in this publication is accurate as of its publication date; such information is subject to change without notice. IEEE is not responsible for any inadvertent errors.

The Institute of Electrical and Electronics Engineers, Incorporated  
3 Park Avenue, New York, NY 10016-5997, USA

Copyright © 2016 by The Institute of Electrical and Electronics Engineers, Incorporated

## Executive Summary

All rights reserved. Published Month 20xx.  
Printed in the United States of America.

IEEE is a registered trademark in the U. S. Patent & Trademark Office, owned by The Institute of Electrical and Electronics Engineers, Incorporated.

PDF: ISBN 978-0-7381-xxxx-x STDVxxxxx

Print: ISBN 978-0-7381-xxxx-x STDPDVxxxxx

IEEE prohibits discrimination, harassment, and bullying. For more information, visit <http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html>.

This work is made available under the [Creative Commons Attribution License](#).

To order IEEE Press Publications, call 1-800-678-IEEE.

Find IEEE standards and standards-related product listings at: <http://standards.ieee.org>

### **Notice and Disclaimer of Liability Concerning the Use of IEEE-SA Industry Connections Documents**

This IEEE Standards Association ("IEEE-SA") Industry Connections publication ("Work") is not a consensus standard document. Specifically, this document is NOT AN IEEE STANDARD. Information contained in this Work has been created by, or obtained from, sources believed to be reliable, and reviewed by members of the IEEE-SA Industry Connections activity that produced this Work. IEEE and the IEEE-SA Industry Connections activity members expressly disclaim all warranties (express, implied, and statutory) related to this Work, including, but not limited to, the warranties of: merchantability; fitness for a particular purpose; non-infringement; quality, accuracy, effectiveness, currency, or completeness of the Work or content within the Work. In addition, IEEE and the IEEE-SA Industry Connections activity members disclaim any and all conditions relating to: results; and workmanlike effort. This IEEE-SA Industry Connections document is supplied "AS IS" and "WITH ALL FAULTS."

Although the IEEE-SA Industry Connections activity members who have created this Work believe that the information and guidance given in this Work serve as an enhancement to users, all persons must rely upon their own skill and judgment when making use of it. IN NO EVENT SHALL IEEE OR IEEE-SA INDUSTRY CONNECTIONS ACTIVITY MEMBERS BE LIABLE FOR ANY ERRORS OR OMISSIONS OR DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR

CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS WORK, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Further, information contained in this Work may be protected by intellectual property rights held by third parties or organizations, and the use of this information may require the user to negotiate with any such rights holders in order to legally acquire the rights to do so, and such rights holders may refuse to grant such rights. Attention is also called to the possibility that implementation of any or all of this Work may require use of subject matter covered by patent rights. By publication of this Work, no position is taken by IEEE with respect to the existence or validity of any patent rights in connection therewith. IEEE is not responsible for identifying patent rights for which a license may be required, or for conducting inquiries into the legal validity or scope of patents claims. Users are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. No commitment to grant licenses under patent rights on a reasonable or non-discriminatory basis has been sought or received from any rights holder. The policies and procedures under which this document was created can be viewed at <http://standards.ieee.org/about/sasb/iccom/>.

This Work is published with the understanding that IEEE and the IEEE-SA Industry Connections activity members are supplying information through this Work, not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought. IEEE is not responsible for the statements and opinions advanced in this Work.