



University of Zanjan  
The Journal of

## **Ethical Reflections**

Vol.1, No.1, Spring 2020, pp. 11-32.

ISSN: 2676-4810

### **Moral Agency in Artificial Intelligence (Robots)**

Saleh Gorbanian<sup>1</sup>

#### **Abstract**

Growing technological advances in intelligent artifacts and bitter experiences of the past have emphasized the need to use and operate ethics in this field. Accordingly, it is vital to discuss the ethical integrity of having intelligent artifacts. Concerning the method of gathering materials, the current study uses library and documentary research followed by attribution style. Moreover, descriptive analysis is employed in order to analyze data. Explaining and criticizing the opposing views in this field and reviewing the related literature, it is concluded that the proof of the ethical integrity of the intelligent artifacts is possible due to the specific features of these artifacts and their own moral system.

**Keywords:** Agency, Ethics, Moral Agency, Intelligent Artifacts, Robot.

---

1. Assistant Professor, Al-Mustafa Open University, salehgh313@gmail.com.  
Received: 11/10/2019 | Accepted: 12/02/2019





دانشگاه زنجان

فصلنامه تأملات اخلاقی

سال اول، شماره اول، بهار ۱۳۹۹، صفحات ۱۱-۳۲.

شاپا: ۴۸۱۰-۲۶۷۶

## عاملیت اخلاقی در مصنوعات هوشمند (ربات‌ها)

صالح قربانیان<sup>۱</sup>

### چکیده

گسترش روزافزون پیشرفت‌های فناوریانه در عرصه مصنوعات هوشمند و تجربه‌های تلخ گذشته از جانب انسان‌های سودجو، لزوم بکارگیری و عملیاتی کردن اخلاق را در این عرصه دوچندان کرده است. در همین راستا، مسئله عاملیت اخلاقی مصنوعات هوشمند مورد بحث و مناقشه است. این پژوهش ابتدا به تبیین دقیق مسئله عاملیت اخلاقی مصنوعات هوشمند (ربات‌ها) و بیان دیدگاه‌های مختلف درباره آنها پرداخته است. سپس با کمک از تحقیقات انجام شده، دیدگاه مخالفان عاملیت اخلاقی مصنوعات هوشمند مورد نقد قرار گرفته است. این تحقیق در روش گردآوری مطالب، کتابخانه‌ای است و در نحوه استناد داده‌ها از شیوه اسنادی پیروی می‌کند و روش آن در تجزیه و تحلیل مطالب، تحلیل محتوایی از نوع توصیفی تحلیلی است. نتیجه این تحقیق آن است که اثبات عاملیت اخلاقی برای مصنوعات هوشمند، با توجه به ویژگی‌های خاص این مصنوعات و نظام اخلاقی مختص به خودشان امکان‌پذیر است.

**واژگان کلیدی:** اخلاق، عاملیت اخلاقی، مصنوعات هوشمند، ربات.

---

تاریخ دریافت: ۹۸/۰۸/۱۹؛ تاریخ پذیرش: ۹۸/۰۹/۱۱

۱. استادیار گروه اخلاق دانشگاه مجازی المصطفی؛ salehgh313@gmail.com

## مقدمه

قرن بیستم و بیست و یک شاهد جهش و پیشرفت‌های انسان در دو عرصه علمی و فناوری بوده است. انسان‌ها در نیمه اول قرن بیستم که به دهه‌های علم معروف است، شاهد پیشرفت در فیزیک و در نیمه دوم آن به جهش‌های علمی در عرصه زیست‌شناسی، ژنتیک و علوم کامپیوتری نائل آمدند. قرن بیست و یک تا به الآن شاهد پیشرفت‌های چشم‌گیر در عرصه‌های عصب‌شناسی، نانو تکنولوژی و رباتیک بوده است. از طرف دیگر قساوت‌هایی که بشر در طول تاریخ از خود به یادگار گذاشته است، گواه این است که اگر انسان امروزی در جهت استفاده درست از این پیشرفت‌های علمی و فناوریانه تربیت نشود، فاجعه‌بارترین دوران را خواهیم داشت. بنابراین، هرچه زودتر باید پاسخ‌های مناسب و قانع‌کننده‌ای برای چالش‌هایی که از طرف این پیشرفت‌های علمی ارائه می‌شود، داده شود. به همین خاطر، نیاز اساسی برای نهادینه و کاربردی کردن اخلاق در عرصه‌های علمی و فناوریانه احساس می‌شود. ما در اینجا با مصنوعات هوشمند سروکار خواهیم داشت و یکی از این مصنوعات هوشمند که به مراتب بیشتر با زندگی ما عجین شده و خواهد شد، ربات‌ها هستند.

پیشرفت‌های فناوریانه شامل سه عرصه زندگی مصنوعی، هوش مصنوعی و رباتیک می‌شوند که نظریه‌های سنتی اخلاق را که به صورت ضمنی یا صریح بر مفهوم شخصیت انسانی استوار بودند، با چالش روبه‌رو کرده است. یکی از چالش‌های بزرگی که ما در اخلاق با آن دست‌به‌گریبان هستیم، مسئله عامل اخلاقی بودن این مصنوعات هوشمند است که روشن شدن جواب این مسئله تأثیرات زیادی در تدوین و تهیه اخلاق کاربردی ماشین‌ها خواهد داشت.

ضرورت پرداختن به مسئله عاملیت اخلاقی مصنوعات هوشمند، از توجه کردن به نقش و اهمیتی است که این موجودات در تشکیل دنیای انسانی دارند. توجه داشتن به این امر باعث اهمیت داشتن چگونگی طراحی و به کارگیری این مصنوعات می‌شود. از طرف دیگر قائل شدن به عاملیت این مصنوعات این امکان را به ما می‌دهد که با جنبه‌های دیگری از ماده (جنبه فیزیکی مصنوعات) توجه داشته باشیم. بدین معنا که وقتی ما عاملیت را برای این مصنوعات قائل شویم، در واقع در حال فکر کردن به این موضوع

هستیم که می‌توان این مصنوعات را به‌عنوان موجودی بی‌روح و بی‌جان، منفعل یا خنثی تلقی کنیم. همه این جواب‌ها نشان از اهمیت داشتن مفهوم عاملیت است. مفهومی که به‌نوعی با یک نوع توانایی و قابلیت انجام عمل (فعل) سروکار دارد. اهمیت مفهوم عاملیت در این است که با مباحث (اخلاقی) مهمی در ارتباط است؛ بامعنای انسان بودن، با مباحث مربوط به روابط انسان‌ها با حیوانات، زمین و موجودات متعالی دیگر، با گفتمان‌های حقوق بشری و با قدرت و مسئولیت‌پذیری.

در رابطه با موضوع عاملیت اخلاقی مصنوعات هوشمند، کتاب‌ها و مقالات متعددی از جانب اندیشمندان مطرح نوشته شده است. مهم‌ترین و مرتبط‌ترین آثار به شرح زیر است: کلبرگ در مقاله‌ی «عاملیت اخلاقی و مسئولیت اخلاقی مجازی» به دنبال پاسخگویی به مجموعه‌ای از سؤالات در مورد مصنوعات هوشمندی مثل ربات‌ها است که سعی می‌کند آن‌ها را به‌عنوان عاملان اخلاقی مسئول معرفی کند. دی بتس در به دنبال اثبات این است که ربات‌ها توانایی این را دارند که اعمال خوب و بد را شناسایی و آن را انجام دهند و همچنین جانسون در مقاله‌ای به نقد نظریات مربوط به فلوریدی و سندرز می‌پردازد که بیشتر ناظر به قدرت تصمیم‌گیری این مصنوعات هوشمند می‌باشد. و اما مقاله حاضر با بهره‌گیری از منابع مختلف، به نقد و بررسی نظرات مخالف و همچنین سعی شده تبیین جامع‌تری از عاملیت اخلاقی مصنوعات هوشمند ارائه دهد.

در ادامه، بعد از مفهوم شناسی عاملیت اخلاقی و پس از بیان چگونگی به‌کارگیری این مفهوم در مورد مصنوعات هوشمند، در ضمن نقد دیدگاه مخالفین، به بیان ملزومات و عدم ملزومات این نوع عاملیت اشاره خواهیم کرد تا در نهایت بتوانیم از ادعای عاملیت اخلاقی داشتن مصنوعات هوشمند دفاع کنیم.

### مفهوم شناسی

اگر بخواهیم عاملیت را با فاعلیت از نظر مفهومی یکسان تلقی کنیم، در این صورت انواع و تقسیمات متنوعی برای فاعلیت می‌توان در نظر گرفت: ۱. بالذات و بالعرض، ۲. قریب و بعید، ۳. خاص و عام، ۴. جزئی و کلی، ۵. بسیط و مرکب، ۶. بالقوه و بالفعل، ۷. ترکیبی از موارد ذکرشده (غفاری، ۱۳۹۴، ص ۳۴۲). اما در فضای اخلاق سکولار، عاملیت روانی یا ذهنی، عاملیت سهمی، عاملیت جمعی، عاملیت ربطی و وابسته و

عاملیت مصنوعی را می‌توان از انواع عاملیت به شمار آورد (Markus, 2019). البته انواع دیگری را هم می‌توان اضافه کرد؛ اما این تقسیم‌بندی بر اساس معیار و نظریات فعلی هست.

الف) عاملیت روانی برگرفته از زندگی روانی یا روحانی است که انسان دارد. انسان همواره در حال توجه، قضاوت، استدلال، ارزیابی، تصمیم‌گیری، قبول کردن و رد کردن و ... است. به نظر می‌رسد این‌ها هم به نوعی فعل به حساب می‌آیند. البته این امر مخالفان و موافقانی هم دارد. ب) عاملیت سهمی زمانی است که دو نفر یا بیشتر کاری را با هم انجام می‌دهند؛ مثل خواندن موسیقی؛ ج) عاملیت جمعی زمانی است که دو نفر یا بیشتر کاری را به عنوان یک گروهی که دارای اصول و ضوابط خاص هست، با هم انجام می‌دهند؛ مثل مؤسسات و ارگان‌های سازمانی. د) عاملیت ربطی و وابسته با توجه به نوعی تحلیل ربطی از خودمختاری و استقلال ناشی می‌شود. فمینیست‌ها قائل هستند تحلیل‌های سنتی از خودمختاری و استقلال، مبتنی بر فردگرایی بوده است و حال آنکه روابط بین افراد در پیشرفت و نگهداری این خودمختاری و استقلال نادیده گرفته شده است. در این دیدگاه استقلال فردی در واقع مانند یک رابطه بین فردی وابسته تحلیل می‌شود. ه) عاملیت مصنوعی در مورد ربات‌ها و سیستم‌های هوشمند مصنوعی هست و اینکه آیا این‌ها توانایی عامل بودن را دارا هستند یا نه؟ ربات‌ها و سیستم‌های مصنوعی با توجه به اینکه فاقد حالت ذهنی و روانی هستند و این امر مخالف با نظریه پذیرفته شده در مورد فعل است که قصد را شرط یک فعل قصدی می‌داند؛ اما آیا با وجود این می‌توان گفت آن‌ها هم به نوعی دارای عاملیت هستند و توانایی انجام برخی کارها را دارند؟ همان‌طور که برخی قائل هستند در پایین‌ترین مرتبه از عاملیت، دارا بودن حالت ذهنی و روانی نیاز نیست.

### انواع عاملیت: ادراکی و گزاره‌ای

ایده دیگری که در عرصه مفهوم عاملیت مطرح شده و به نوعی مفهوم عاملیت را گسترش داده است، تقسیم عاملیت به عاملیت گزاره‌ای و عاملیت ادراکی است. باید عرض کنم که فلسفه عمل (فعل) مبتنی بر یک ایده ساده است: «تفاوت داشتن فعل با عمل». این

تفاوت یک نوع تفاوت شهودی است؛ مثل فرق بلند کردن دست با تپش قلب که در اولی ارادی و دومی غیرارادی است، نیست. بنابراین، فلاسفه عمل دنبال این امر بودند که این تفاوت شهودی بین فعل با عمل را توضیح دهند. به روش‌های مختلفی می‌شود به این سؤال جواب داد. عمل به کاری گفته می‌شود که از حیوان از روی قصد انجام داده باشد و اخص از فعل است؛ زیرا فعل گاهی به کاری که حیوانات بدون قصد آن را انجام می‌دهند، اطلاق می‌شود و بعضاً به فعل جمادات نیز اطلاق می‌شود؛ ولی عمل خیلی کم در مورد فعل بدون قصد و فعل جمادات بکار می‌رود (راغب اصفهانی، ۱۴۱۲، ج ۱، ص ۵۸۷). عمل به کاری گفته می‌شود که از روی عقل و فکر باشد؛ لذا با علم مقرون است، ولی فعل اعم از آن است؛ گاه ممکن است توأم با تعقل و تفکر باشد و گاه بدون تعقل و تفکر انجام پذیرد. پس رابطه بین عمل و فعل رابطه اعم و اخص است. این فرق را از قرآن مجید نیز می‌شود استفاده کرد که عمل به اعمال و کارهای ارادی اطلاق شده؛ ولی فعل گاهی در افعال جمادات نیز بکار رفته است؛ مثل آیه ۶۳ سوره انبیاء: «بَلْ فَعَلَهُ كَبِيرُهُمْ هَذَا» که در این آیه «فعل» به جمادات نسبت داده شده است؛ زیرا آیه مذکور مربوط به داستان حضرت ابراهیم (ع) است که (چون تبر به عنوان وسیله جرم نزد «بت» بزرگ بود) فرمود از «بت» بزرگ سوال کنید. بنا براین بت که از جمله جمادات است «فعل» به او منتسب شده است (قرشی، ۱۴۱۲، ج ۵، ص ۴۵). برخی از اهل لغت گفته‌اند: عمل عبارت است از آنچه در خارج واقع شده است و کاملاً به مرحله تحقق رسیده است؛ اما فعل عبارت است از صدور عمل همراه با اختیار و قصد؛ و این عمل ممکن است از انسان، حیوان، جن و شیطان صادر شود و نیز ممکن است صالح باشد یا غیر صالح که دارای اثر طبیعی، جزایی و الهی است (مصطفوی، ۱۳۶۸، ج ۸، ص ۲۲۵). نظر مولف تاج‌العروس این است که مابین فعل و عمل از چند جهت فرق است: اول اینکه عمل به کاری گفته می‌شود که با تفکر و تدبیر انجام گیرد و فعل هر نوع کاری را شامل می‌شود، خواه همراه با فکر و تدبیر باشد و خواه نباشد. دوم اینکه «عمل» اشاره به استمرار کار دارد، ولی در فعل استمرار شرط نیست و به همین جهت خداوند در قرآن «واعملوا صالحاً» فرموده است نه «وافعلوا صالحاً». سوم اینکه «عمل» فقط کارهای اختیاری را در بر می‌گیرد؛ ولی «فعل» کارهای اختیاری و غیر اختیاری را شامل می‌شود. چهارم اینکه

«عمل» قول (سخن) را در بر دارد، ولی «فعل» این گونه نیست (حسینی زبیدی، ۱۴۱۴، ج ۳۰، ص ۵۵ و ۱۸۲). اما فلاسفه عمل هم به این تفاوت توجه داشته‌اند و از آنجایی که در موقع طرح این سؤال نظر به ذهن داشته‌اند به همین سبب موقع جواب دادن هم با توجه به این امر جواب داده‌اند. به همین دلیل بیشتر فلاسفه درگیر این قضیه شده‌اند که کدام یک از عاملیت‌ها ناشی از عمل بر طبق دلیل است. برای روشن شدن جواب این سؤالات باید به تفاوت‌های میان عاملیت گزاره‌ای و عاملیت ادراکی توجه داشت.

عاملیت گزاره‌ای همان دیدگاه سنتی در مورد عاملیت است، در این نوع عاملیت توانایی انجام عمل بر طبق قضاوت‌های دلیل‌دار مورد بررسی قرار می‌گیرد. بدین معنا که اگر شما بخواهید ساندویچی را بخورید با این استدلال که انرژی و قوت کافی برای انجام کارهای روزمره‌تان را داشته باشید و سپس به این دلیل هم ساندویچ را خورده باشید، در واقع شما یک عامل گزاره‌ای به حساب می‌آید (Sebo, 2017, p. 3). بنابراین بدون داشتن توانایی تفکر گزاره‌ای نمی‌توانیم در مورد افعال، باورها و امیالمان فکر کنیم. در بین دانشمندان اختلاف نظر وجود دارد که در چه محدوده‌هایی می‌توانیم تفکر گزاره‌ای داشته باشیم و بر اساس آن عمل کنیم. «ژوسه برمودز» تفکر گزاره‌ای را نیازمند زبان گزاره‌ای می‌داند و اینکه تنها عده معدودی از انسان‌ها هستند که دارای زبان گزاره‌ای هستند. در نتیجه، تنها عده معدودی از انسان‌ها دارای تفکر گزاره‌ای هستند (Sebo, 2017, p. 4). بنابراین، غیر انسان نمی‌تواند عمل گزاره‌ای داشته باشد و حتی انسان‌های معلول هم این گونه خواهند بود.

از طرف دیگر برخی باور دارند عاملیت گزاره‌ای دارای دامنه عریضی است. «دوروتی چنی» و «رابرت سیفارس» بر این باور هستند که تفکر گزاره‌ای نیازمند زبان گزاره‌ای نیست و درست برعکس آن را قبول دارند و آن اینکه انسان‌ها از زبان عمومی برای تعامل با دیگران استفاده می‌کنند (Sebo, 2017, p.5). نتیجه این دیدگاه این می‌شود که بسیاری از حیوانات و انسان‌ها می‌توانند فکر کنند و عمل گزاره‌ای داشته باشند، بدون اینکه بتوانند صحبت گزاره‌ای داشته باشند. این دانشمندان حتی استدلال می‌کنند که نوعی بوزینه بنام «بابون» دارای قدرت زبان تفکری است.

فرقی ندارد که قائل به کدام یک از این دیدگاه‌ها (محدود یا گسترده) باشید، چیزی



که مهم است این است که عاملیت گزاره‌ای خیلی از رفتارهای خلاقانه و هدف‌دار را که در انسان‌ها و حیوانات قابل مشاهده است، نمی‌تواند توجیه کند؛ آن‌هم به سه دلیل: یک) با وجود اینکه می‌توان گفت برخی از انسان‌ها دارای توانایی عمل گزاره‌ای هستند، ولی نمی‌توان گفت تمامی انسان‌ها دارای این توانایی هستند. دو) همواره این گونه نیست که تمامی انسان‌ها توانایی عمل گزاره‌ای را داشته باشند؛ بلکه انسان‌ها این توانایی را به تدریج کسب می‌کنند. سه) بر فرض داشتن چنین توانایی، این گونه نیست که ما همواره از این توانایی مان استفاده کنیم؛ بلکه در برخی مواقع طبق عمل گزاره‌ای رفتار می‌کنیم؛ مثل زمان‌هایی که به مشکل برمی‌خوریم و در سایر موقعیت‌ها صرفاً از روی احساس اینکه این امر درست است عمل می‌کنیم؛ یعنی از روی احساس و طبیعت. اگر این دلایل درست باشد، با وجود وسیع بودن عاملیت گزاره‌ای، باز تمایل داریم بدانیم فرق عاملیت بین فعل و رفتار چیست؟ پس نیاز به تبیین نوع دیگری از عاملیت احساس می‌شود. اینجا است که بحث از عاملیت ادراکی به میان می‌آید.

عاملیت ادراکی: این نوع عاملیت بیشتر در بین فلسفه قاره‌ای قابل مشاهده است تا در فلسفه تحلیلی. دلیل این امر هم شاید این باشد که شواهد اولیه‌ای که برای این نوع از عاملیت وجود دارد، پدیداری است. پدیده‌هایی مثل حافظه، انتظارات، باورها، امیال و سایر اشکال روان‌شناسانه ما که باعث تولید تجربه‌های ادراکی ما می‌شوند. پس عاملیت ادراکی به توانایی ما برای عمل کردن بر طبق تجربه‌های ادراکی هنجاری مان برمی‌گردد. این نوع عاملیت شامل توانایی ارزیابی در مورد کارهایی که انجام می‌دهیم نیز می‌شود. قابلیت دیگر عاملیت ادراکی عمل بر طبق شهود است. این توانایی شامل بسیاری از حیوانات و انسان‌ها می‌شود و حال آنکه عاملیت گزاره‌ای این گونه نبود. بسیاری از انسان‌ها دارای هردو نوع عاملیت (گزاره‌ای و ادراکی) هستند؛ بچه قبل از حرف زدن دارای عاملیت ادراکی است و پس از رسیدن به عاملیت گزاره‌ای دلیلی برای از دست دادن اولی وجود ندارد؛ به دلیل اینکه عاملیت ادراکی به مراتب سودمندتر و قابل استفاده‌تر است و حتی بیشتر هم مورد استفاده قرار می‌گیرد. دلیل دیگر، تجربه شخصی افراد از داشتن و عمل کردن بر طبق عاملیت ادراکی است. ما کارهای روزمره زیادی انجام می‌دهیم بدون اینکه از خودمان برای انجامشان دلیل بخواهیم و یا مرحله به مرحله از

خودمان بیرسیم الآن چه کار باید کرد.

این تفاوت مهم بین عاملیت ادراکی و عاملیت گزاره‌ای باعث به وجود آمدن یک سؤال هنجاری بسیار مهم می‌شود و آن اینکه چه حقوق و مسئولیت‌هایی برای عاملیت ادراکی لازم و کافی است؟ و اینکه اگر انسان‌ها و غیر انسان‌ها دارای عاملیت یکسان هستند، آیا دارای شأن اخلاقی یکسان هم هستند؟ جواب مثبت است و به همین دلیل است که بیشتر فیلسوفان گستره‌عامل اخلاقی را فراتر از انسان‌ها می‌دانند و حیوانات، مصنوعات هوشمند و غیره را هم جزو عاملان اخلاقی می‌شمارند.

مفهوم عاملیت اخلاقی بدین معنا است که موجودی توانایی انجام افعال اخلاقی را داشته باشد. نکته مهم دیگری که باید به آن اشاره کرد این است که تبیین عاملیت اخلاقی ارتباط بسیار وثیقی با تحلیل ما از اخلاق و اخلاقیات دارد؛ اما به‌طور کلی دو مورد توانایی وجود دارد که برای عاملیت اخلاقی ضروری هستند: شرط اولی که معرکه آراء متضاده است، توانایی انتخاب آزادانه برای عمل کردن هست. ملاک ارزش افعال انسانی است، اختیار است (مصباح یزدی، ۱۳۹۲، ص ۱۰۷). شرط دوم مربوط به قوه عقل می‌شود و آن توانایی فهم و درک درستی و نادرستی است. کسی که فرق بین درست و نادرست را تشخیص نمی‌دهد، در قبال رفتارهایش هم مسئول نخواهد بود. توانایی اول با دو شرط در ارتباط است. لازمه انتخاب آزادانه عمل این است که عامل، توانایی سنجش و تعمق را داشته باشد و از طرف دیگر، عاقل نیز باشد. منظور از تمیز دادن بین خوب و بد، علم چگونگی به کارگیری مفاهیم و اصول اخلاقی است و لازمه مشترک این دو توانایی اصلی، برخوردار بودن از آگاهی و خودآگاهی است (De Sio, 2016, p. 2).

### دیدگاه مخالفان عاملیت اخلاقی مصنوعات هوشمند (ربات‌ها)

عاملیت اخلاقی داشتن مصنوعات هوشمند، درست به‌مانند سایر نظریات چالشی دیگر، مخالفان و موافقانی دارد. چهار دلیل مخالفان که غیرانسانها را (مثل ربات‌ها، مصنوعات هوشمند و ...) عامل اخلاقی نمی‌دانند، عبارتند از: یک) اعتراض غائی: عامل‌های مصنوعی نمی‌توانند هدف داشته باشند؛ دو) اعتراض قصدی: عامل‌های مصنوعی حالت قصدی نمی‌توانند داشته باشند؛ سه) اعتراض آزادی: عامل‌های مصنوعی آزادی ندارند؛

چهار) اعتراض مسئولیت‌پذیری: عامل‌های مصنوعی نمی‌توانند در قبال اعمالشان مسئول باشند. اگر بخواهیم این اعتراضات را در قالب یک مفهوم کلی‌تر ارائه کنیم، باید به سراغ مفهوم «شخصیت» برویم. به عبارت دیگر، مخالفان، عامل اخلاقی را به شخصی که دارای یک سری قابلیت‌ها و توانایی‌های استدلالی است، تعریف می‌کنند. بنابراین، می‌شود سؤال کرد که آیا فرضی که براساس آن این افراد «شخص» را تعریف می‌کنند، درست است؟ دوم اینکه آیا فناوری را هم می‌توان «شخص» فرض کرد و آن را عامل اخلاقی به حساب آورد؟ (Luppincini, 2008, p. 9). استدلال کامل این گروه به این صورت است که عامل اخلاقی شخصی است که دارای آزادی اراده و حد معقولی از ذهن است که یکی از ثمرات این ذهن خودآگاهی داشتن است.

از منظرهای مختلفی می‌توان مصنوعات هوشمند را به‌عنوان عامل‌های اخلاقی فرض کرد. منظر اول این است که این مصنوعات باید چه نوع ذهن و مغزی داشته باشند تا به‌عنوان عامل اخلاقی به حساب آیند. این راه‌کار از جانب «فلوریدی» و «سندرز» پیشنهاد شده است و با تقلیل و کاهش در چارچوب‌های عامل اخلاقی، مصنوعات را هم جزو عامل‌های اخلاقی محسوب می‌کنند. منظر دوم مربوط به این است که این مصنوعات چه کاری برای ما انجام می‌دهند. این پیشنهاد هم از جانب «وربیک» مطرح شده است که عامل اخلاقی را به‌واسطه کاری که انجام می‌دهد تعریف می‌کند. منظر سوم از جانب «مگ‌نینی» مطرح شده است و آن تعریف عامل اخلاقی از طریق ارزش‌هایی است که دارند (Coeckelbergh, 2009, p. 1)؛ به عبارت دیگر می‌توان گفت که مصنوعات هوشمند چه ارزش‌هایی دارند تا بدین‌وسیله در حیطه عامل‌های اخلاقی وارد شوند.

بعضی‌ها فراتر از این هم رفته‌اند و ادعا کرده‌اند که انسان در بهترین حالت به‌عنوان یک عامل اخلاقی معیوب و ناقص به حساب می‌آید و عامل‌های اخلاقی مصنوعی تنها عامل‌هایی هستند که بهترین مدل برای عامل اخلاقی بودن باشند. افرادی که از این بیان حمایت و طرفداری می‌کنند «پیتر دنیلسون» و «هنس مریوق» هستند. یکی از استدلال‌های قائلان به عامل اخلاقی بودن مصنوعات این‌گونه است:

مقدمه اول: به‌طور کلی، نوزادان و کودکان کوچک به‌اندازه کافی توسعه‌یافته نیستند

تا از نظر اخلاقی مسئولیت اعمالشان را داشته باشند. مقدمه دوم: هنگامی که به سن بلوغ می‌رسیم، به لحاظ اجتماعی بالغ می‌شویم و می‌توانیم نوعی روابط بین فردی با یکدیگر داشته باشیم که شامل اخلاقی بودن یکدیگر از طریق تحسین یا سرزنش اخلاقی است. مقدمه سوم: بزرگسالانی که مسئولیت اخلاقی دارند، قادر به انجام اعمال اخلاقی، به‌عنوان عامل اخلاقی هستند (Cross, 2015, p. 9). نتیجه‌ای که از این سه مقدمه گرفته شده است این است که مسئولیت و عامل اخلاقی سازه‌ای تکاملی است.

از این استدلال به نفع مصنوعات هوشمند هم استفاده شده است. بدین صورت که عوامل هوش مصنوعی ممکن است در آینده به نقطه‌ای از توسعه دست یابند تا جایی که وقتی ما احساسات خود را در برابر اقداماتشان نشان می‌دهیم، می‌توانند ستایش (قدردانی) و یا سرزنش (خشم) را از خودشان نشان دهند. نتیجه اینکه از آنجایی که ما احساس قدردانی یا خشم نسبت به اعمال عوامل مصنوعی داریم، بنابراین می‌توان گفت که آن‌ها از نظر اخلاقی مسئول و عامل اخلاقی به حساب می‌آیند.

### هستی‌شناسی مصنوعات هوشمند

در ابتدا بهتر است که جایگاه این مصنوعات در هستی‌روشن شود. برای تمایز و شناسایی موجوداتی که در جهان وجود دارند، از مباحث مربوط به هستی‌شناسی کمک گرفته می‌شود. از منظر هستی‌شناسی شاید بتوان از سه وجود اساسی و بنیادی یاد کرد: طبیعی، انسانی و مصنوعی. البته این سبک تقسیم‌بندی هستی‌شناسانه با توجه به مفهوم عاملیت است. انسان‌ها هم بر مصنوعات و هم بر طبیعات عاملیت دارند و قدرت دخل و تصرف در این‌ها دارند. هر کدام از موجودات طبیعی و مصنوعی هم عاملیت مختص به خودشان را دارا هستند (Johnson & Miller, 2008, p. 1). در اثر پیشرفت جوامع و فناوری‌های جدید، مصنوعات بشری وارد عرصه‌های مختلفی از جوامع بشری شده‌اند و زندگی انسان‌ها با این مصنوعات عجین شده است. از این رو بود که سؤال بعدی مطرح شد؛ اگر این موجودات (مصنوعی) دارای عاملیت هستند، چرا عاملیت آن‌ها اخلاقی نباشد و اینکه چرا آن‌ها نباید دارای عاملیت اخلاقی باشند؟

بنابراین افکار و اندیشه‌های مربوط به هستی‌شناسی مصنوعات، عاملیت مصنوعی و

عامل اخلاقی بودن مصنوعات باهم دیگر در آمیختند. برای اینکه بحث از عامل اخلاقی بودن مصنوعات واضح تر شود، بعد از هستی‌شناسی مصنوعات باید چگونگی عامل بودن این مصنوعات روشن گردد. در مورد هستی‌شناسی مصنوعات چنین گفته می‌شود: مصنوعات معمولاً به اشیاء ساخته شده توسط انسان اطلاق می‌شود. مصنوعات دارای طبیعت دوگانه‌ای هستند. مصنوعات فنی که توسط مهندسين برای اهداف عملی طراحی می‌شوند، و از یک سو دارای جنبه‌های فیزیکی هستند که دارای ویژگی‌های هندسی، فیزیکی و شیمیایی‌اند و از طرف دیگر دارای جنبه کاربردی هستند که وظیفه خاصی را دارا هستند و از این لحاظ با حالات ذهنی و اعمال قصدی ارتباط تنگاتنگی دارند (Houkes, 2006, p. 2). این طبیعت دوگانه که جزو مسائل اساسی در هستی‌شناسی مصنوعات به حساب می‌آیند. جنبه فیزیکی مصنوعات جزء حقایق طبیعی به حساب می‌آید و مستقل از گرفتاری‌ها و مسائل انسانی هستند؛ اما جنبه کاربردی مصنوعات خاصیت نهادی و سازمانی دارد (Pohjola, 2007, p. 14). این جنبه نهادی و سازمانی مصنوعات با درگیری‌های انسانی در ارتباط است و مستقل از آن‌ها نیست.

در ادامه، مباحث را در دو مرحله پیش خواهیم گرفت تا از این طریق هم به تبیین استدلال موافقان عاملیت اخلاقی برای مصنوعات هوشمند پرداخته و هم اینکه به نقدها و ایرادهای مخالفین این امر جواب داده باشیم. برای رسیدن به این منظور بحث را در دو قالب «ملزومات موردنیاز برای عامل اخلاقی بودن مصنوعات» و «معیارهایی که مورد نیاز عامل اخلاقی مصنوعی نیست» پیش خواهیم برد.

### ملزومات موردنیاز برای عامل اخلاقی بودن مصنوعات

برای اینکه یک تصمیم و عملی به صورت تصادفی و در موقعیت‌های خاص، خوب تلقی شود، چه ملزوماتی لازم دارد؟ چهار عنصر اساسی در کنار هم نیازهای اساسی برای اخلاقی بودن را تشکیل می‌دهند: تجسم، یادگیری، همدلی و هدف‌گایی. این فهرست از ملزومات را می‌توان با سایر نظرها در مورد عامل اخلاقی بودن مصنوعات مقایسه کرد؛ به‌ویژه که با نظریات فلوریدی و سندرز مشترکات زیادی دارد که قائل به تعامل، استقلال و خودمختاری و سازگاری هستند (DeBaets, 2014, p. 3)؛ ولی باید توجه داشت که مابین نظر اولی و دومی هم تفاوت خاص‌تگاهی و منشائی وجود دارد و هم

اینکه هر کدام دارای مزایای خاص به خود هستند.

استدلال اول در مورد تجسم خواهد بود و اینکه عامل اخلاقی مصنوعی هم درست به مانند تمامی عامل های اخلاقی دیگر باید نوعی تجسم یافتگی را دارا باشند؛ چراکه آنها باید در دنیای فیزیکی (نه مجازی) تأثیر گذاری داشته باشند؛ البته در صورتی که بخواهند اخلاقی عمل کنند (DeBaets, 2014, p. 4). بنابراین، تجسم یافتگی به آنها اجازه خواهد داد که با عمل اخلاقی مورد نظر ارتباط واقعی داشته باشند. همچنین، با آثار و پیامدهای مربوط به عملشان می توانند قدرت انطباق و یا پاسخ گویی داشته باشند و این تجسم یافتگی لزوماً به معنای شباهت صد درصدی داشتن با جسمانیت انسانها و عملشان نیست.

یادگیری هم صرفاً به معنای جمع آوری یک سری اطلاعات نیست، بلکه به معنای تعدیل و سازگار کردن تصمیمات اخلاقی و اعمالش بر پایه اطلاعات قبلی است. ویژگی سازگاری و تعدیل به عامل این امکان را می دهد که همواره از اشتباهاتش به مانند موفقیت هایش یاد بگیرد و استدلال های اخلاقی اش را توسعه دهد و اطلاعات جدید به دست آمده از محیط را در تصمیماتش اعمال کند. ویژگی یادگیری باعث ارتقاء تصمیم گیری عامل در وضعیت های جدید خواهد شد، آن هم با توجه به آثار مثبت و منفی که ناشی از تصمیمات گذشته اش بوده است و با توجه به این آثار و نتایج مثبت و منفی حاصل از تصمیمات گذشته، در آینده تصمیمات بهتری خواهند گرفت (دی بیتس، ص ۴). در شکل ساده اش، عامل اخلاقی مصنوعی از آثار و نتایج اعمال گذشته اش در جهت پیروی و متابعت بهتر از قوانینی که بر پایه این قوانین طراحی شده است، یاد می گیرد؛ اما در شکل پیشرفته، عامل اخلاقی مصنوعی از این توانایی یادگیری برای پیشرفت و توسعه دادن (به اصطلاح طراحی و برنامه ریزی دوباره) این قوانین پایه ای که بر اساس آنها طراحی شده، استفاده می کند. این مدل پیشرفته می تواند مدلی برای توسعه فضیلت های اخلاقی باشد؛ بدین معنا که علاوه بر چیزهایی که به عامل گفته شده که اینها خوب هستند، عامل می تواند حسها و رفتارهای اخلاقی و قضاوتی خود را در پاسخ به آنچه آموخته است، تحقق و توسعه بخشد.

اما باز هم با وجود اینکه یک ربات تجسم یافته می تواند از اعمال گذشته اش یاد بگیرد،

الزاماً به معنای اخلاقی بودن این ربات نیست. او باید توانایی همدلی هم داشته باشد؛ اما چرا همدلی برای اخلاقی بودن لازم است؟ دلیل اول؛ همدلی به عامل اجازه می‌دهد تا زمانی که با یک عامل اخلاقی دیگر و یا حتی با یک استدلال اخلاقی مناسب روبه‌رو شد، آن را تشخیص دهد. (قدرت تشخیص عامل اخلاقی دیگر و همچنین استدلال‌های اخلاقی مناسب). این حس همدلی به عامل اخلاقی مصنوعی این توانایی را می‌دهد که از نیازها و امیال دیگران و ضرر و سودهایی که به دیگران وارد می‌شود، آگاهی داشته باشد و آن‌ها را بفهمد و درک کند. البته این امر وابسته به نظریه ذهن است. به این معنا که شناخت این که یک موجود دیگری با افکار، باورها، ارزش‌ها و نیازهای خاص به خودش وجود دارد، مبتنی بر وجود ذهن است (DeBaets, 2014, p. 4). البته اینجا هم دارا بودن ذهن برای داشتن رفتارهای اخلاقی موردنیاز نیست؛ بلکه همین مقدار که عامل بتواند ارزش‌های موجودات دیگر را مورد ارزیابی قرار دهد کافی است و این امر از طریق حس‌گرهایی که به زبان، استرس، تپ صدا و سایر نشانه‌ها حساس هستند می‌تواند صورت گیرد و از طریق ارزیابی این موارد می‌تواند تشخیص دهد که عامل مقابل (دیگری) آیا اخلاقی عمل می‌کند یا نه؟ با استفاده از این حس‌گرها عامل مصنوعی می‌تواند به صداهای بلند عکس‌العمل نشان دهد، ولی باید این قدرت تشخیص را هم مابین صداهایی بلندی که برای درخواست کمک هستند و غیر آن را داشته باشد. همدلی برای توسعه و ارزیابی انتخاب‌های اخلاقی ضروری است؛ همان‌طور که احساسات برای انسان عاقل ضروری است، بنابراین همدلی هم برای ماشین‌های اخلاقی هم ضروری است.

در تبیین هر نوع عامل اخلاقی (حتی عامل اخلاقی مصنوعی هم) باید گرایش و تمایل داشتن به سوی اخلاقی بودن، مدنظر قرار گیرد و از این امر نباید غافل بود. بدین معنا که همان‌طور که انسان‌ها بعضی خصوصیات و حس‌های اخلاقی را در خودشان رشد می‌دهند و تمام تلاش خود را می‌کنند که براساس آن عمل کنند، عامل‌های مصنوعی اخلاقی هم باید به گونه‌ای طراحی شوند که به دنبال خوبی و خیر باشند. باین وجود، این عامل‌های مصنوعی اخلاقی در صورت عدم وجود برخی طراحی‌ها در این زمینه، این کار را انجام نخواهند داد. البته برای ربات‌ها لازم نیست که براساس

نظریات پیچیده اخلاقی طراحی شده باشند، به گونه‌ای که انسان‌ها در حین انجام تصمیمات اخلاقی شان به کار می‌برند؛ بلکه این مصنوعات می‌توانند برای این امر برنامه‌ریزی شوند.

### معیارهای غیر لازم برای عامل اخلاقی

در ادامه مطلب گذشته که بحث از ملزومات مورد نیاز برای عامل اخلاقی مصنوعی بود، در اینجا به مواردی اشاره خواهد شد که برای عامل اخلاقی بودن و توسعه این عاملیت ضروری نیستند. اولین و شایع‌ترین معیاری که در بین عموم مردم و در طول تاریخ از آن بسیار یاد شده است، هوشی است که به معنای عقلانیت می‌باشد. هوش در عاملیت اخلاقی کاربرد زیادی ندارد. برخی از دانشمندان پیش‌بینی می‌کنند که در آینده قدرت محاسباتی به قدری پیشرفت خواهد کرد که در نهایت منجر به هوش مافوق انسانی خواهد شد که نتیجه‌اش عاملیت برتر هم خواهد بود (DeBaets, 2014, p. 5). ولی به نظر می‌رسد که این گونه نیست. درست است که یک حد خاصی از هوش و قدرت محاسبه برای عامل بودن ضروری است که آن هم جنبه کاربردی دارد، چرا که باعث سهولت در امر یادگیری و همدلی می‌شود؛ اما داشتن قدرت پردازش قوی باعث نمی‌شود که یک عاملی دارای تفکر بالا باشد و همچنین است دارا بودن هوش بالا، به خودی خود باعث بیشتر اخلاقی بودن نمی‌شود. البته عامل اخلاقی بودن نیاز به یک حداقل سطحی از هوش دارد برای اینکه عامل بتواند انتخاب‌هایش را انجام دهد و بر طبق آن‌ها عمل کند، اما وجود آن‌ها به عنوان عوامل اخلاقی، عمدتاً یک نوع کیفیت‌هایی هستند که در تعامل با جهان و هستی به وجود می‌آیند و همچنین از تجربیاتی که کسب می‌کنند، یاد می‌گیرند. مورد دیگر که برای عامل اخلاقی بودن ضروری نیست، «تبعیت از قانون» است. در حالی که یک حد خاصی از این امر در مورد عامل‌های مصنوعی ضروری است، ولی مقدار زیاد آن منجر به نوعی برده‌داری برنامه‌ریزی شده خواهد انجامید (DeBaets, 2014, p. 5). عاملیت اخلاقی نیازمند قدرت تصمیم‌گیری و عمل بر طبق آن تصمیم است؛ آن‌هم زمانی که مابین برخی از قواعد اساسی تناقض رخ می‌دهد؛ و همچنین دست‌شستن از قوانین و عدم پیروی از آن‌ها زمانی که اجبار و اکراه حاکم است (موقعیت



جبری). گفته می‌شود تا زمانی که عاملی توانایی انتخاب بد بودن را نداشته باشد، خوب بودن او بی‌معنا خواهد بود؛ بنابراین صرف تبعیت از یک قانون و عمل نکردن به آن در صورت تناقض، باعث عامل اخلاقی بودن نمی‌شود.

یک ماشین باید بتواند فراتر از صرف پیروی کردن از قوانین عمل کند تا بتواند به‌عنوان عامل اخلاقی تلقی شود، اما از طرف دیگر داشتن آزادی کامل برای عامل اخلاقی بودن ضروری نیست. باید یک ماشین دارای سازگاری، انعطاف‌پذیری، برنامه‌ریزی و اولویت‌بندی باشد، اما در دنیای مادی آزادی‌ها همدیگر را مقید می‌کنند (یک آزادی دیگر را محدود و مقید می‌کند)؛ بنابراین داشتن آزادی تام و کمال برای عامل اخلاقی بودن لازم و ضروری نیست. تمام آزادی‌ها توسط محیط و فهم ما از آن آزادی و همچنین توسط اعمالمان و اجاب‌های بیرونی مقید می‌شوند. همچنین از جانب اعمال گذشته‌مان، تعداد گزینه‌های موجود برای انتخاب کردن و حتی توسط توانایی‌هایمان در کاربست انتخابمان، نیز به‌نوعی در محدودیت هستیم و آزادی‌مان مقید می‌شود؛ بنابراین در مورد ماشین‌ها هم، همین‌گونه است که طراحی، هدف خاصی داشتن از این طراحی، دانش‌های پیشینی به ماشین‌ها داده می‌شود و حتی توسط محیط محدود می‌شوند.

بعضی دیگر معتقد هستند که داشتن یک نوع آگاهی کاملاً انسانی، برای پیشرفت و توسعه عامل اخلاقی بودن لازم است؛ اما این مورد هم می‌تواند در ماشین‌ها به‌صورت کاملاً متفاوت از نوع انسانی آن وجود داشته باشد (DeBaets, 2014, p. 6). مفهوم «آگاهی» بدون درک فرآیند آن، به‌راحتی قابل توضیح نیست. در مورد خودآگاهی می‌گویند: آگاهی از اعمال خود، یا اعمال خود را متعلق به خود دانستن؛ یا آگاه بودن از خود به‌عنوان یک فرد (Marine Webster's Online Dictionary). آگاهی نسبت به خود، آگاهی به ظهورات و اقدامات دیگران؛ آگاهی از وجود خود، به‌ویژه دانش خود را به‌عنوان یک موجود آگاه دریافتن (Oxford Online Dictionary). اغلب تصورشان بر این است که یک نوع آگاهی در تجربه و احساس کردن لازم است. به‌عنوان مثال داشتن درد به معنای احساس کردن درد است و برای احساس درد باید آگاهی به درد وجود داشته باشد. همچنین است در مورد یک شیء؛ برای اینکه آن را با

چشم سر بینید باید آن شیئی قابل رؤیت تجربی باشد، برای رؤیت تجربی شیئی باید به آن آگاهی داشته باشید. در مورد باورها هم این گونه است، می‌توانند از روی آگاهی یا به صورت ناخود آگاه باشند و اما آن‌هایی که به صورت آگاهانه و با توجه به باورهای دیگر شکل گرفته باشند، منطقی و معقول به نظر می‌رسند (Tye, 1996, p. 2).

به راستی آگاهی چیست؟ ممکن است سؤال راحتی به نظر برسد، اما این گونه نیست. آگاهی هم مفهوم واضح و روشنی است و هم بسیار سخت و مشکل. ما از آگاهی برای شناخت و بررسی خودمان استفاده می‌کنیم. سختی و پیچیدگی این مفهوم باعث شده بود که سال‌های متمادی فیلسوفان و دانشمندان از مطالعه و تحقیق درباره این مفهوم روی گردان باشند، اما خوشبختانه در آغاز قرن بیست و یکم بود که مطالعات در مورد آگاهی رونق یافت. چون در این دوران روانشناسی، زیست‌شناسی و علوم اعصاب به مرحله‌ای رسیده بودند که آمادگی پاسخگویی به برخی سؤالات پیچیده را داشتند.

البته این امر به معنای زدوده شدن رمز و راز از مفهوم آگاهی و شعور نبود، ما الآن فقط اطلاعات زیادی در مورد ذهن داریم و اگر از این نقطه هم بخواهیم به سوی شناخت آگاهی قدم برداریم با مشکلات بسیاری روبه‌رو هستیم. البته برخی ادعا دارند که پرده از راز و رمز آگاهی کنار زده‌اند، به نظر می‌رسد که آن‌ها یک‌جانبه به مسأله نگاه کرده‌اند و از شکاف عمیقی که مابین دنیای فیزیکی و روانی وجود دارد، غافل شده‌اند. (مسئله ذهن و بدن) مشکلی که فلاسفه برای بیش از دو هزار سال با آن دست‌به‌گریبان هستند. مشکل این است که در تجربه متداول انسانی، به نظر می‌رسد دو چیز کاملاً متفاوت وجود دارد و ما هیچ روش آشکار و واضحی برای تبیین این دو در کنار یکدیگر نداریم. از یک طرف، تجربه‌های شخصی ما وجود دارد؛ دیدن اشیاء که در اطراف ما هستند و لذت بردن و یا حتی ناراحت شدن از آن‌ها. همه این‌ها تجربه‌های شخصی من هستند و کیفیتی دارند که نمی‌توانم آن‌ها را به هیچ کس دیگری منتقل کنم. آیا تجربیات شما در مورد رنگ سبز درست مثل تجربه من از رنگ سبز است، یا اینکه بویی که از یک غذای خوشمزه استشمام می‌کنیم یکی است، این‌ها مسائلی هستند که ما هرگز نمی‌توانیم بفهمیم.

از سوی دیگر، معتقد هستیم که جهان فیزیکی‌ای وجود دارد که این تجربیات را به وجود می‌آورد. ممکن است در مورد آنچه از آن ساخته شده‌اند، یا در مورد ماهیت

واقعی‌شان تردید داشته باشیم، اما شک نداریم که آن‌ها وجود دارند. مشکل این است که این دو نوع چیز (اشیاء و جهان فیزیکی و تجربیات شخصی) کاملاً متفاوت هستند. چیزهای فیزیکی واقعی وجود دارند که دارای ابعاد، شکل، وزن و سایر ویژگی‌ها هستند که هر کس می‌تواند آن‌ها را اندازه‌گیری کند و در موردشان به توافق برسد، اما از طرف دیگر تجربیات خصوصی وجود دارند، مثل احساس درد، رنگ‌ها و غیره که در موردشان توافق وجود ندارد (Blackmore, 2017, p. 14).

در طول تاریخ انسان‌ها برای حل این مشکل، قائل به وجود دو دنیا شدند: دنیای مادی و دنیای روحانی. ادیان اصلی تقریباً همه دوگانه نگر هستند. در فرهنگ‌های غربی، حتی در میان افراد غیرمذهبی هم دوگانگی شایع است. این دوگانه‌انگاری به قدری قوی است که حتی در زبان روزمره ما هم نفوذ کرده و افراد موقع صحبت واژه‌هایی مثل «بدن من»، «ذهن من» را به کار می‌برند که به نوعی حکایت از دوگانگی می‌کند. در قرن هفدهم، فیلسوف فرانسوی رنه دکارت به طور رسمی تئوری دوگانه‌انگاری را ارائه کرد. در دوگانگی دکارتی، ذهن و مغز از مواد مختلف تشکیل شده‌اند. با توجه به نظر دکارت، ذهن غیر فیزیکی و غیرمطول است (فضا یا موقعیتی ندارد)، درحالی که بدن از مواد فیزیکی و دارای خاصیت گسترش تشکیل یافته است؛ اما سؤال این است که چگونه این دو باهم تعامل برقرار می‌کنند؟ دکارت جواب می‌دهد که آن‌ها از طریق غده کوچک صنوبری شکل که در مرکز مغز قرار دارد، باهم در ارتباط هستند (Blackmore, 2017, p. 15). این غده صنوبری شکل دارای ساختار فیزیکی است و دوگانگی دکارتی هیچ توضیحی نمی‌دهد، مبنی بر اینکه چرا این غده به تنهایی نمی‌تواند با قلمرو ذهنی ارتباط برقرار کند.

بنابراین بنظر می‌رسد این مفهوم هنوز جای بحث دارد و به راحتی نمیتوان آن را به عنوان معیاری ضروری برای عاملیت اخلاقی مصنوعات هوشمند بکار برد. تردید و سؤال از آگاهی و ضرورت آن برای عاملیت اخلاقی با سایر موارد بحث شده، به کل متفاوت می‌باشد. برخی از فیلسوفان از مسئله آگاهی برای تردید در عامل اخلاقی بودن ماشین‌ها سود جسته‌اند و قائل به یک چیز مبهمی (روح داشتن) شده‌اند. (آگاهی مستلزم دارا بودن روح است). مسئله روح داشتن درست به مثابه مفهوم خدا برای انسان است، انسانی که می‌خواهد اخلاقی باشد، آیا

نیازی به مفهوم خدا دارد یا نه؟ جواب احتمالی است و ضرورتی وجود ندارد بر اینکه انسان بخواهد یک موجود اخلاقی باشد، حتماً مفهوم خدا هم وجود داشته باشد. البته باید توجه داشت که کسی مزایای مفهوم خدا را برای اخلاقی بودن آن هم در برخی مواضع به هیچ وجه رد نمی‌کند؛ بلکه بحث از ضرورت این مفهوم است.

### نتیجه‌گیری

عاملیت اخلاقی یک مفهوم هنجاری است و دسته‌ای از موجودات را که رفتارهایشان موضوع برای الزامات اخلاقی هست را شامل می‌شود. به عبارت دیگر عاملیت اخلاقی یعنی توانایی داشتن برای اینکه یک موجود اخلاقی به حساب آید. عاملیت اخلاقی یعنی داشتن توانایی تصمیم‌گیری و به شیوه اخلاقی مسئولانه عمل کردن را گویند. عاملیت اخلاقی از عناصر زیر ترکیب شده است؛ توانایی عمل به شیوه فضیلت‌مندانه، داشتن مهارت‌های فکری نظیر تفکر منطقی، قصد و ذهن که این امکان را فراهم می‌کند تا صاحب این ویژگی‌ها قدرت درک قوانین اخلاقی را داشته باشد و درنهایت توانایی پذیرفتن مسئولیت عمل‌هایش و اینکه دارای شأنی باشد که دیگران در قبال اعمالشان نسبت به او پاسخ‌گو باشند. بنابراین با توجه مطالب گذشته میتوان عاملیت اخلاقی را برای ربات‌های هوشمند اثبات کرد.

نکته آخری که باید مورد توجه قرار گیرد این است که عاملیت اخلاقی در قالب و بطن یک سیستم اخلاقی تعیین می‌گردد. منظور این است که برای تعیین و ترسیم عامل اخلاقی، در حله اول باید نظام اخلاقی مورد نظر ترسیم شده باشد، در مرحله بعد نوبت به این امر می‌رسد که در این چنین نظام اخلاقی چه موجودی و با چه معیارهایی می‌تواند عامل اخلاقی تلقی شود. بنابراین در اثبات عاملیت اخلاقی مصنوعات هوشمند هم نباید آنها را با سایر عامل‌های اخلاقی مثل انسان یکی پنداشت.

## منابع

- حسینی زبیدی، محمد مرتضی، (۱۴۱۴ق)، تاج العروس من جواهر القاموس، محقق / مصحح: علی هلالی و علی سیری، بیروت: دارالفکر، چاپ اول.
- راغب اصفهانی، حسین بن محمد، (۱۴۱۲ق)، المفردات فی غریب القرآن، تحقیق: صفوان عدنان داودی، دمشق و بیروت: دارالعلم و الدار الشامیه.
- غفاری ابوالحسن، (۱۳۹۴)، قانون علیت در فلسفه اسلامی و تحلیل تطورات، قم: سازمان انتشارات پژوهشگاه فرهنگ و اندیشه اسلامی چاپ اول.
- قرشی بنایی، علی اکبر، (۱۴۱۲ق)، قاموس قرآن، تهران: دار الکتب الاسلامیه، چاپ ششم.
- مصباح یزدی، محمد تقی، (۱۳۹۲)، انسان‌شناسی در قرآن، تنظیم و تدوین محمود فتحعلی، قم: انتشارات مؤسسه آموزشی و پژوهشی امام خمینی.
- مصطفوی، حسن (۱۳۶۸)، التحقیق فی کلمات القرآن الکریم، تهران، چاپ اول.

Blackmore, S. (2017). *Consciousness: A very short introduction*. Oxford University Press.

Coeckelbergh, M. (2009). *Virtual moral agency, virtual moral responsibility: on the moral significance of the appearance, perception, and performance of artificial agents*. *AI & society*, 24(2), 181-189.

Cross, G. E. (2015). "Can morality be Computed? An Exploration of whether Machines Can be Moral Agents" (Doctoral dissertation; Available at; <http://search.ndltd.org>).

De Sio, F. S. & van Wynsberghe, A. (2016). "When should we use care robots? The nature-of-activities approach". *Science and Engineering Ethics*, 22(6), 1745-1760.

DeBaets, A. M. (2014). "Can a robot pursue the good? Exploring artificial moral agency". *Journal of Evolution and Technology*, 24(3), 76-86.

DeBaets, A. M. (2014). "Can a Robot Pursue the Good? Exploring

- Artificial Moral Agency". *The Journal of Evolution and Technology*, 24(3), 76-86.
- Houkes, W. & Meijers, A. (2006). "The ontology of artefacts: the hard problem". *Studies in history and philosophy of science*, Part A, 37(1), 118-131.
- Johnson, D. G. & Miller, K. W. (2008). "Un-making artificial moral agents". *Ethics and Information Technology*, 10(2-3), 123-133
- Johnson, D. G. & Noorman, M. (2014). "Artefactual agency and artefactual moral agency". In: *The Moral Status of Technical Artefacts* (pp. 143-158). Springer, Dordrecht.
- Luppicini, R. (Ed). (2008). *Handbook of research on technoethics*. IGI Global.
- Nagenborg, M. (2007). "Artificial moral agents: an intercultural perspective". *International Review of Information Ethics*, 7(09), 129-133.
- Peterson, M. & Spahn, A. (2011). "Can technological artefacts be moral agents?" *Science and Engineering Ethics*, 17(3), 411-424.
- Pohjola, P. (2007). *Technical artefacts: an ontological investigation of technology*. Jyväskylä: University of Jyväskylä.
- Sebo, J. (2017). "Agency and moral status". *Journal of Moral Philosophy*, 14(1), 1-22.
- Tye, M. (1996). "The Function of Consciousness". *Noûs*, 30 (3), 287-305. Available at:
- Marine Webster's Online Dictionary. "Self-consciousness". Available at: <https://www.merriam-webster.com/dictionary/self-consciousness>.
- Oxford Online Dictionary. "Self-consciousness". Available at: <https://en.oxforddictionaries.com/definition/self-consciousness>.