



بخش ۷:

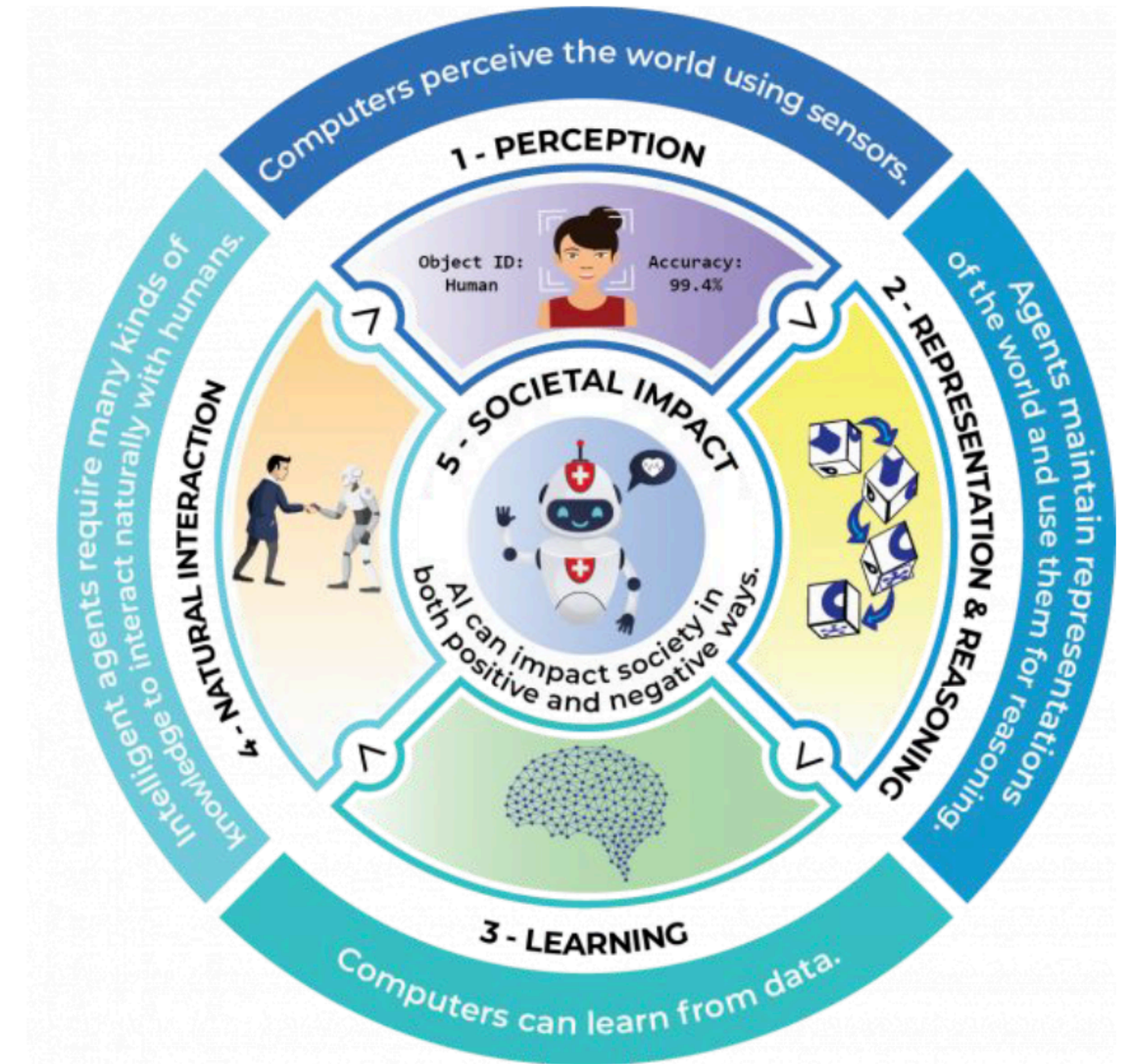
اخلاق و استفاده مسئولانه
از هوش مصنوعی



ایده‌های کلیدی هوش مصنوعی

اثرگذاری اجتماعی پنجمین «ایده کلیدی هوش مصنوعی» است.

ما امیدواریم که ابزارهای هوش مصنوعی که ایجاد می‌کنیم منصفانه و صادقانه مورد استفاده قرار گیرند - به عبارت دیگر مطمئن باشیم که آنها تأثیر منفی بر جامعه ندارند. این نوع کاربرد را استفاده "اخلاقی" یا مسئولانه از یک ابزار می‌نامیم.





نقش هوش مصنوعی در زندگی روزمره

👉 هوش مصنوعی در حوزه‌های مختلفی از زندگی روزمره تأثیرگذار خواهد بود، از جمله:

- **بهداشت و درمان:** تشخیص بیماریها، مدیریت پرونده‌های پزشکی و ارائه مشاوره‌های سلامت
- **آموزش:** یادگیری شخصی‌سازی‌شده، ارزیابی خودکار و تولید محتوای آموزشی
- **خدمات مالی:** تحلیل داده‌های مالی، پیش‌بینی بازار و ارائه مشاوره‌های سرمایه‌گذاری
- **حمل‌ونقل:** خودروهای خودران، مدیریت ترافیک و بهینه‌سازی مسیرها و موارد متعدد دیگر



دیدگاه مغرضانه

اگر یک دیدگاه مورد توجه قرار گیرد یا نادید گرفته شود چه اتفاقی می‌افتد؟

در مدل ساده اخلاقی ما بر اساس سود و زیان مشکلی وجود دارد. چه می‌شود اگر ما به سادگی برخی از دیدگاه‌ها را نسبت به دیدگاه‌های دیگر ترجیح دهیم یا نادیده بگیریم؟

این منجر به سوگیری می‌شود

"سوگیری عبارت است از ترجیح یا تعصب نسبت به یک یا چند گروه، خصوصاً به شیوه‌ای که ناعادلانه یا غیرواقعی تلقی شود."



درک سوگیری داده‌ها

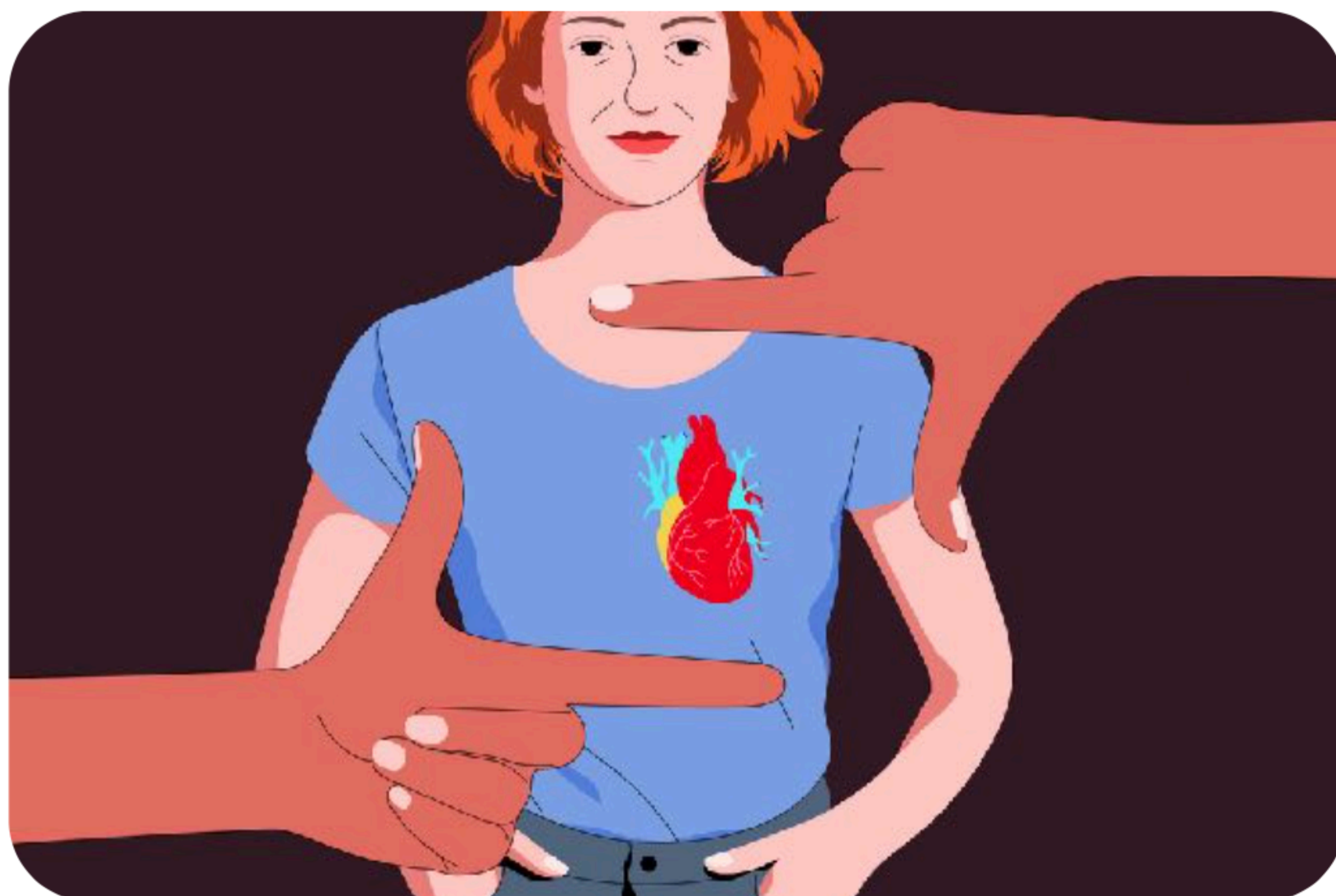
وقتی داده‌های یادگیری ماشین ما **سوگیرانه** باشد، اوضاع بدتر می‌شود. اگر انسان‌ها **متعصب** شوند، می‌توانند با دریافت اطلاعات جدید نظر خود را تغییر دهند و از آن درس بگیرند. اما هوش مصنوعی به این راحتی نمی‌تواند این کار را انجام دهد.

هوش مصنوعی زمانی که داده‌های مورد استفاده برای آموزش هوش مصنوعی **مغرضانه** باشد، می‌تواند **سوگیرانه** شود. ما به این **سوگیری داده** می‌گوییم. همیشه واضح نیست که آیا داده‌ها در مرحله اول سوگیری بوده‌اند یا خیر، و اگر قرار به تغییر باشد هوش مصنوعی باید با حجم بسیار بالای داده‌های جدید **دوباره آموزش داده شود** تا سوگیری را حذف کند.



یک مثال از سوگیری داده

چرا یک مجموعه داده بزرگ، منصفانه و فراگیر مهم است



یک سیستم هوش مصنوعی را در نظر بگیرید که بیماری قلبی را پایش می کند.

به طور تجربی هم مردان و هم زنان علائم مشترکی مانند درد قفسه سینه را تجربه می کنند، اما زنان ممکن است سوء هاضمه، تنگی نفس و کمردرد را نیز تجربه کنند.



یک مثال از سوگیری داده

چرا یک مجموعه داده بزرگ، منصفانه و فراگیر مهم است



اگر یک سیستم هوش مصنوعی فقط با استفاده از مجموعه داده‌ای از مردان مبتلا به بیماری قلبی آموزش داده شود، ممکن است درد قفسه سینه را به عنوان علائم بیماری قلبی تشخیص ندهد، و در نتیجه بازدهی آن برای خانم‌ها همراه با خطای بسیار خواهد بود.



یک مدل ساده اخلاق

یکی از معانی اخلاق یعنی درک درست و غلط

ما قصد داریم یک مدل ساده از اخلاق بسازیم تا تصمیم بگیریم که آیا یک سیستم هوش مصنوعی خاص باید استفاده شود یا خیر. بلوک های سازنده این سیستم دو روشی است که هوش مصنوعی می تواند بر افراد تأثیر بگذارد:

👉 **زیان** چیزی است که برای فرد یا گروهی از افراد اثر منفی داشته باشد.

👉 **سود** چیزی است که نتیجه مثبتی برای یک فرد یا گروهی از افراد دارد.

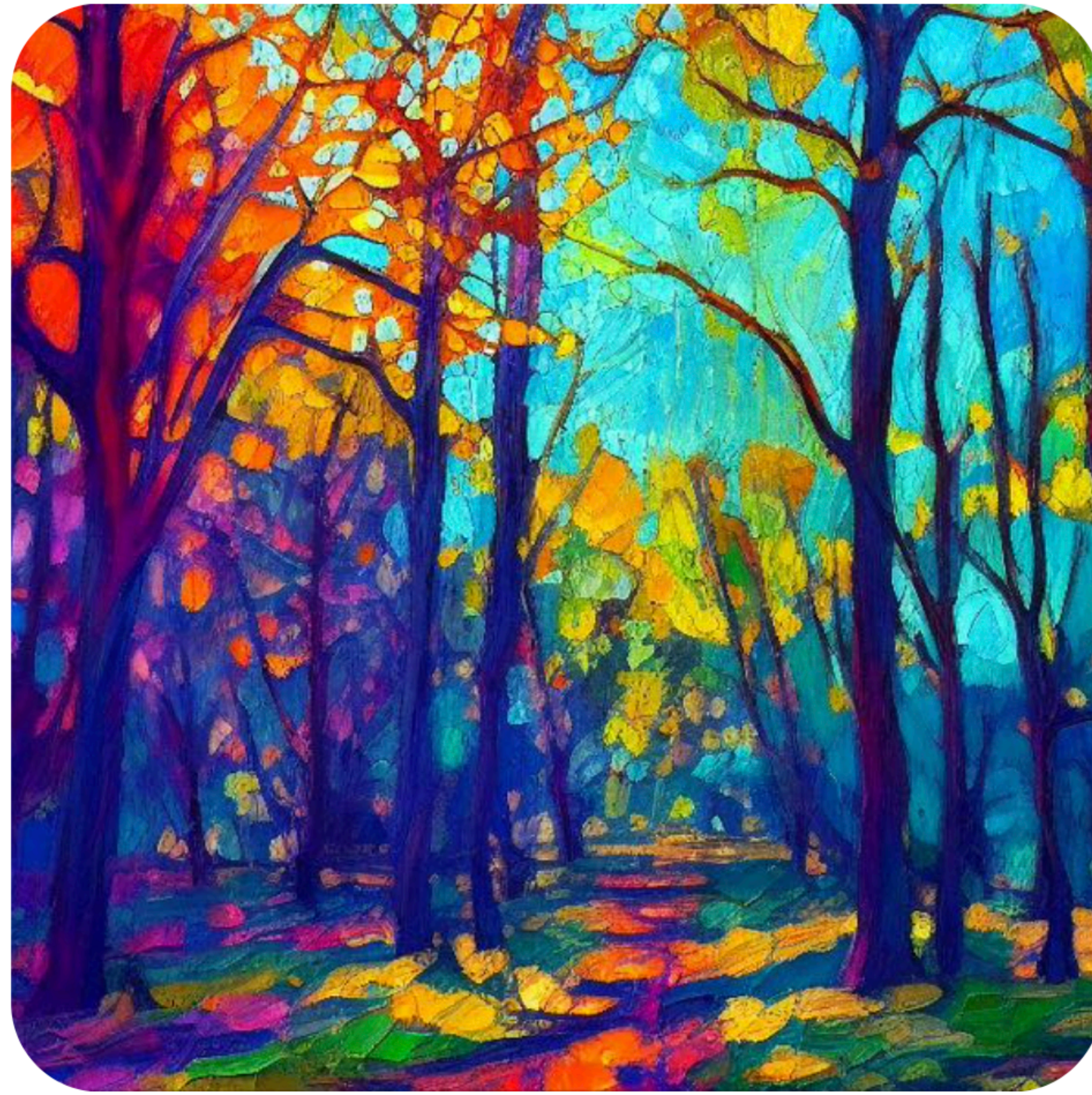
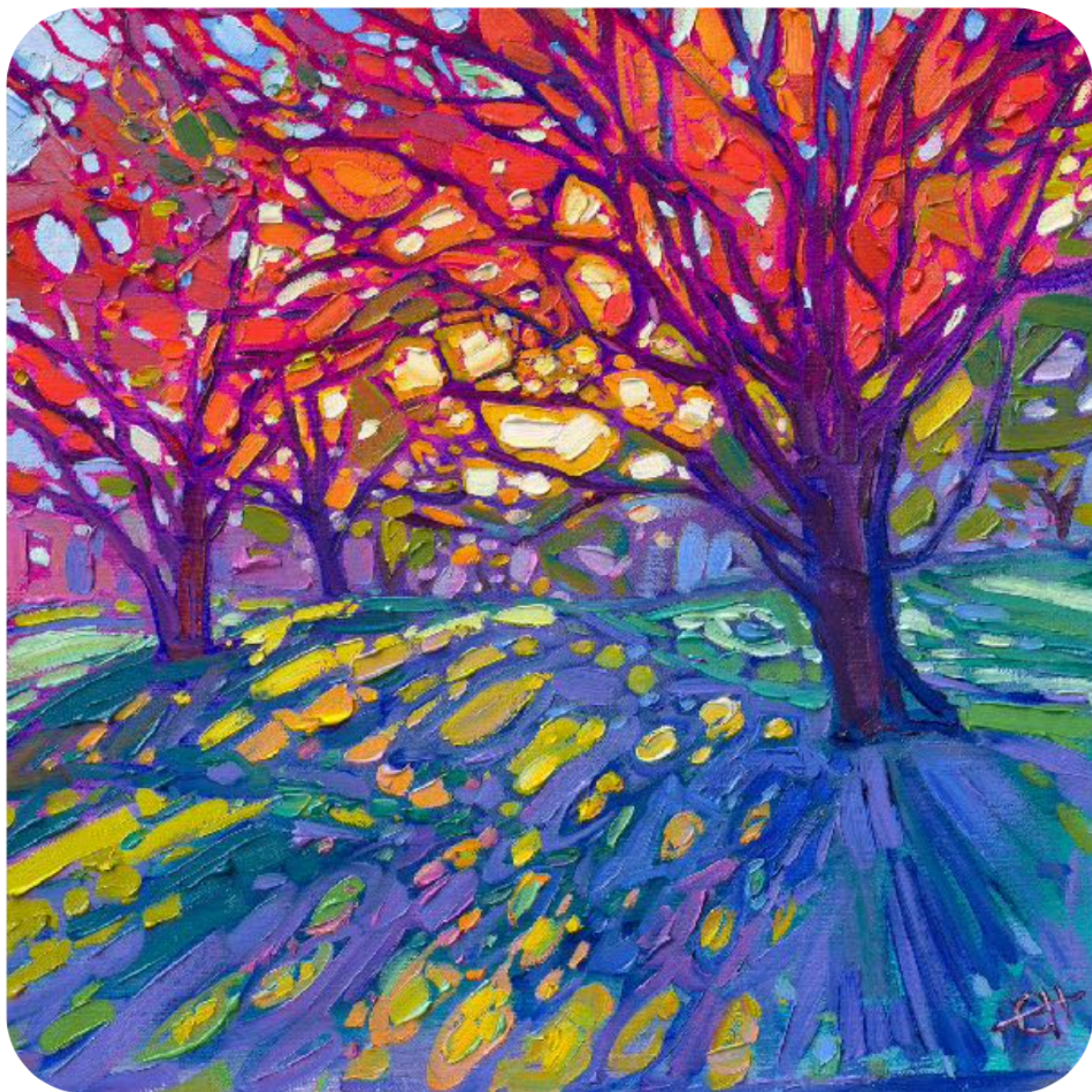


به این تصویر نگاه کنید





این یکی یک تصویر مشابه است





درختان افرا کریستالی،
ارین هانسون ۱۴۰۰



اثر هنری ساخته شده با
Stable Diffuion AI





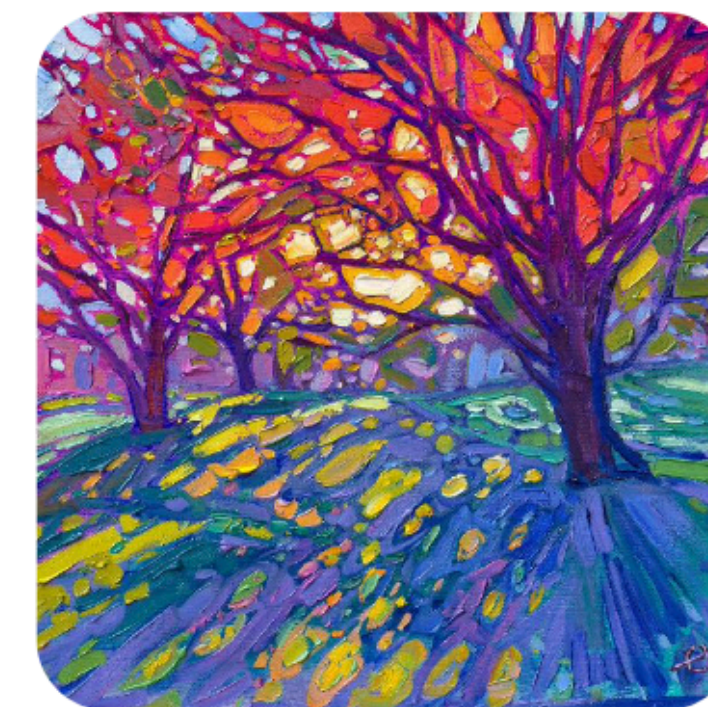
درک سود و زیان

سرقت ادبی/هنری: برداشتن کار یا ایده های دیگران و به اشتراک گذاشتن آنها و نسبت دادن به خود.

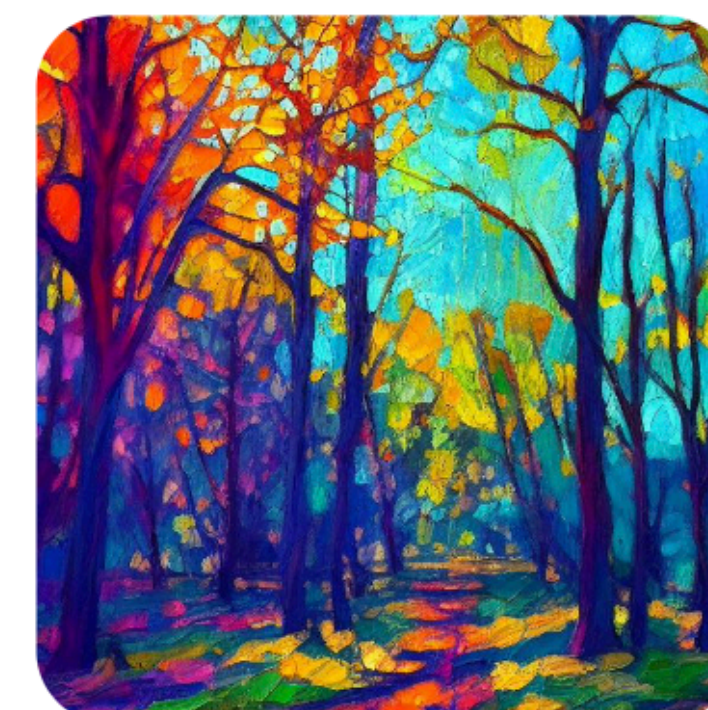
هنر هوش مصنوعی مانند آنچه با Stable Diffusion ساخته شده است باید با استفاده از آثار هنری واقعی ساخته شده توسط انسان آموزش داده شود.

👉 آیا فکر می کنید هنرهای تولید شده با هوش مصنوعی مانند این **سرقت ادبی/هنری** است؟ چرا و چرا نه؟

👉 اگر هنرمند **بداند** که از کارش برای آموزش هوش مصنوعی استفاده می شود، پاسخ شما تغییر می کند؟



اثر خلق شده توسط انسان



ساخته شده با هوش مصنوعی



بررسی جامعه ذینفعان

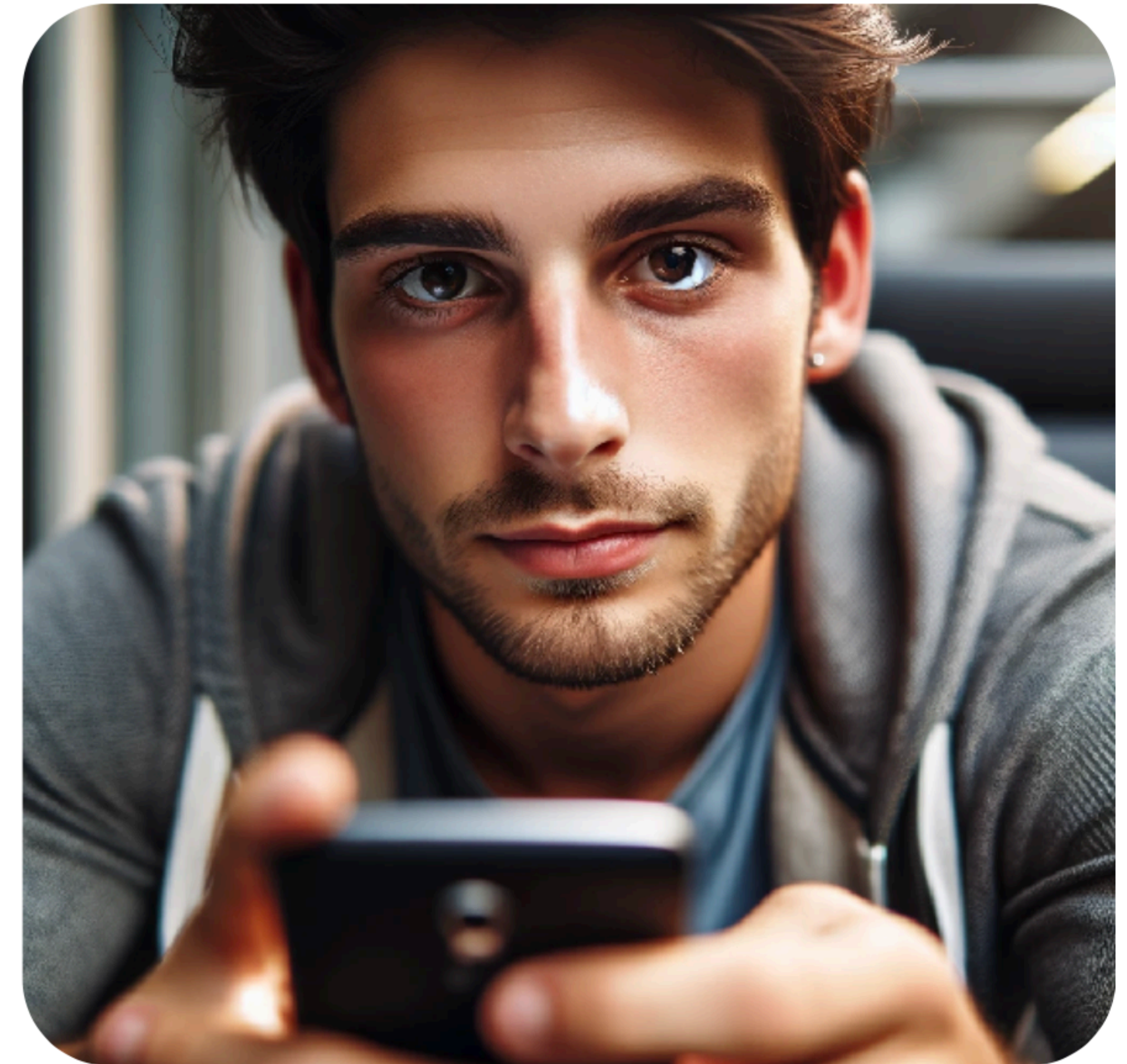
جامعه خبری



خواننده

"من به عنوان مصرف کننده اخبار، نمی‌دانم دقیقاً درباره استفاده از هوش مصنوعی در روزنامه نگاری چه حسی داشته باشم. این امر تحویل اخبار را سرعت می بخشد و آن را مطابق با علایق ما سفارشی می‌کند، اما نگرانی‌هایی را نیز به همراه دارد. هوش مصنوعی ممکن است حاوی اطلاعات سوگیرانه را از سوی سازندگان خود باشد و به انتشار سریع اخبار جعلی دامن بزند و به اعتماد به اخبار و دموکراسی آسیب برساند. از طرفی ارائه اطلاعات انتخابی می‌تواند حقیقت را تحریف کند.

من نسبت به بدتر شدن مسائلی مانند انتشار اخبار جعلی، شکاف اجتماعی، و تضعیف اعتبار روزنامه نگاری محتاط هستم. برای پرداختن به این موضوع، به دنبال منابع خبری شفاف هستم که اطلاعات خود را به دقت تأیید می‌کنند و از هوش مصنوعی مسئولانه استفاده می‌کنند. ما برای اطمینان از اینکه هوش مصنوعی از حقیقت و دموکراسی حمایت می‌کند، به قوانین روشن و باز نیاز داریم.»



تصاویر با هوش مصنوعی DALLE ساخته شده و متن‌ها بوسیله ChatGPT نوشته شده



سیاستمدار

«به عنوان یک سیاستمدار، موضوع استفاده از هوش مصنوعی در روزنامه‌نگاری نگرانی‌های اخلاقی سختی را به همراه دارد. هوش مصنوعی می‌تواند به تجزیه و تحلیل داده‌ها کمک کند و افراد بیشتری را به اخبار علاقه‌مند کند، اما همچنین می‌تواند سوگیری‌ها را بدتر کند و آنچه را که مردم می‌بینند دستکاری کند. این ممکن است باعث شود مردم کمتر به اخبار اعتماد کنند.»

انتشار سریع اطلاعات جعلی یا "تغییر شده" می‌تواند به دموکراسی و اینکه چقدر خوب با هم کنار می‌آییم آسیب می‌رساند. به همین دلیل است که به عنوان یک سیاستمدار در انظار عمومی، بر **اطمینان از شفاف بودن الگوریتم‌ها** تمرکز می‌کنم، به مردم یاد می‌دهم اخبار را بهتر درک کنند و قوانین محکمی برای حفظ صداقت روزنامه‌نگاری و ایمن نگه داشتن دموکراسی در عصر اینترنت ایجاد می‌کنم.»



تصاویر با هوش مصنوعی DALLE ساخته شده و متن‌ها بوسیله ChatGPT نوشته شده



خبرنگار

"به عنوان یک خبرنگار، استفاده از هوش مصنوعی در کار من فرصت‌هایی را ممکن می‌سازد اما همچنین سوالات اخلاقی را مطرح می‌کند. هوش مصنوعی می‌تواند محتوایی مانند مقاله‌های خبری و وبلاگ‌ها را شخصی‌تر و سریع‌تر کند، اما همچنین سوالاتی در مورد حریم خصوصی، الگوریتم‌های ناعادلانه و انتشار اخبار جعلی ایجاد می‌کند.

واقعاً مهم است که مطمئن شوید الگوریتم‌ها واضح هستند، داده‌های افراد ایمن نگه داشته می‌شوند و صداها متنوعی در یادگیری هوش مصنوعی گنجانده شده است. ما به افرادی نیاز داریم که فناوری را توسعه دهند تا با دولت، اخلاق مداران و دیگران همکاری کنند تا قوانین محکمی وضع کنند و مطمئن شوند که همه از آنها پیروی می‌کنند.

با مدیریت خوب این چالش‌ها، می‌توانیم از هوش مصنوعی برای تغییر روزنامه‌نگاری به سمت بهتر استفاده کنیم و آن را صادقانه و مورد اعتماد مردم نگه داریم.»





چارچوب‌ها و توافق‌نامه‌های بین‌المللی در زمینه اخلاق هوش مصنوعی

۱. توصیه‌نامه اخلاق هوش مصنوعی یونسکو (UNESCO AI Ethics Recommendation)

در سال ۲۰۲۱، یونسکو نخستین استاندارد جهانی در زمینه اخلاق هوش مصنوعی را تصویب کرد که برای تمامی ۱۹۴ کشور عضو قابل اجراست. این توصیه‌نامه بر اصولی مانند:

- تناسب و جلوگیری از آسیب
- ایمنی و امنیت
- حفظ حریم خصوصی و حفاظت از داده‌ها
- حاکمیت چندذینفعی و تطبیقی
- مسئولیت‌پذیری و پاسخگویی
- شفافیت و قابلیت توضیح



چارچوبها و توافقنامه‌های بین‌المللی در زمینه اخلاق هوش مصنوعی

۲. اصول اخلاقی سازمان ملل متحد برای استفاده از هوش مصنوعی

سیستم سازمان ملل متحد مجموعه‌ای از ده اصل را برای راهنمایی استفاده اخلاقی از هوش مصنوعی در تمامی مراحل چرخه عمر سیستم‌های هوش مصنوعی ارائه داده است. این اصول بر پایه اخلاق و حقوق بشر بنا شده‌اند و شامل مواردی مانند: حفظ کرامت انسانی | پاسخگویی و شفافیت | عدم تبعیض | حفاظت از حریم خصوصی

۳. کنوانسیون شورای اروپا درباره هوش مصنوعی، حقوق بشر، دموکراسی و حاکمیت قانون

در ۲۰۲۴، شورای اروپا نخستین معاهده الزام‌آور بین‌المللی در زمینه هوش مصنوعی را تصویب کرد. این کنوانسیون با هدف تضمین استفاده مسئولانه از هوش مصنوعی و احترام به حقوق بشر، دموکراسی و حاکمیت قانون تدوین شده است. ۴۶ کشور عضو شورای اروپا به همراه ۱۱ کشور غیرعضو، از جمله ایالات متحده، کانادا، ژاپن و اسرائیل، در تدوین این معاهده مشارکت داشته‌اند.



AlgorithmWatch is a non-governmental, non-profit organization based in Berlin and Zurich. We fight for a world where algorithms and Artificial Intelligence (AI) do not weaken justice, human rights, democracy, and sustainability but strengthen them.

