

اخلاق و علم داده

مایک لوکیدز، هیلاری میسن و دی.جی پاتیل

Ethics and Data science

Mike Loukids, Hilary Mason & DJ Patil

اطلاعات کتابنامه ای :

اخلاق و علم داده

مایک لوکیدز، هیلاری میسن و دی.جی پاتیل

Copyright © 2018 o’rielly Media. All rights reserved

.Printed in the united states of America

...

Illustrator : Rebecca Demarest

....

پیشگفتار

در سالهای اخیر شاهد افزایش چشمگیر توجه همگانی به موضوع اخلاق داده بوده ایم. علت این توجه، اخبار و داده های جعلی و اتفاقاتی از این دست [در انتخابات 2016] می باشد. البته نگرانی ها درباره اخلاق داده مدتها قبل از انتخابات 2016¹ آغاز شده بود. خیلی قبل تر از زمانی که سیستم تگ گذاری عکس گوگل به اشتباه برخی افراد سیاه پوست را به عنوان گوریل شناسایی کرد؛ این نگرانی ها از زمانی که برای اولین بار بحث علم داده مطرح شد -و شاید حتی پیش از آن- شروع شد و هر روز بیشتر و بیشتر میشود.

دلیل این توجه را می توان در به کارگیری داده ها در تمام ابعاد زندگی مان جستجو کرد: درخواستهای دوستی یا ارتباطات کاری که دریافت می کنیم، لیست کالاهای شامل تخفیف که در ایمیل مان دریافت می کنیم، اخباری که می بینم، آهنگهایی که گوش می دهیم. درواقع با هر حرکتی که انجام می دهیم داده های ما جمع آوری می شود، هر حرکتی که حاکی از حضور آنلاین و یا حتی گاهی حضور فیزیکی ماست. هرچند استفاده درست از داده، در زندگی مزایای زیادی برای ما داشته است، اما گاهی سوء استفاده از آن موجب رنجش و آزار ما شده است. بسیاری از این نگرانی ها در گزارشهایی که کاخ سفید درباره داده و هوش مصنوعی تهیه کرده است مورد تاکید قرار گرفته اند. برای مثال، ابلاغیه به کلیه ایالات آمریکا برای الزام به گنجاندن اخلاق داده و امنیت در تمام برنامه های آموزشی علم داده و تکنولوژی.

دیدن گردهمایی افراد در رویدادهایی مثل D4G² و FAT³ برای تبادل نظر در مورد اخلاق و مشاهده بحث های پرشور آن ها در مورد اصول اخلاقی در **کانال اسلک Data For Democracy** و روی آوردن وبلاگ نویسان و منتقدان به موضوع اخلاق، بسیار عالی و در خور تقدیر است.

اما حلقه گمشده در این میان، یافتن روشی برای به کار بستن اخلاق در استفاده از داده و به طور کلی در فرایند توسعه محصول است. تعهد اخلاقی صرفا پذیرفتن یک مجموعه از اصول نیست، بلکه تغییر نحوه عملکرد شماست. مثلا اینکه بگویم قبل از استفاده از داده های کاربران در یک آزمایش، باید از آنها اجازه گرفت یک بحث است و عمل به آن در ابعاد وب یک بحث دیگر است و در عین حال اینکه در یک وب سایت، به طور شفاف توضیح بدهیم که از داده های کاربران قرار است چطور استفاده شود و این استفاده به چه نتایجی منجر خواهد شد و بعد اجازه و رضایت کاربران برای این استفاده را بدست بیاوریم، یک بحث کاملا متفاوت است. این چیزی است که نیاز به بررسی آن داریم.

درک این نکته هم ضروری است که اخلاق، یک لیست ثابت از باید ها و نباید ها نیست، بلکه بحثی است درباره اینکه آنچه شما انجام می دهید چه اثراتی بر دیگران خواهد داشت و آیا این اثرات قابل قبول هستند یا نه.

این چیزی است که موضوع بحث این کتاب است: عملی کردن اخلاق. و عملی کردن اخلاق به معنای ایجاد فضای بحث، ایجاد فضای اختلاف نظر و حصول اطمینان در مورد بررسی پیامدها و نتایج یک پروژه در هر مرحله از آن و بسیاری موضوعات دیگر است. اخلاق در زمینه استخدام چه می گوید؟ در یک محیط آکادمیک اخلاق را چطور باید آموزش دهیم؟ همه اینها سوالات بزرگی هستند که نمی توان در کتاب کوتاهی مثل این کتاب به آنها پاسخ داد. با اینحال لازم است در مورد این سوالات صحبت کنیم.

¹ انتخابات سال 2016 ریاست جمهوری آمریکا که طی آن در شبکه های اجتماعی از روشهای مشکوکی برای تبلیغ و اقناع رای-دهندگان برای انتخاب کاندیداها استفاده شد.

² Data For Good Exchange 2019

³ ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*)

علم داده یک ورزش تیمی است و ما شما را در این تیم لازم داریم. با توجه به سرعت تکنولوژی و تحول فکری ای که در مورد داده انتظار داریم، این کتاب را یک پروژه تکرار شونده می دانیم. مثل نسخه های محصولات فناورانه که از نسخه 0.1 به نسخه 1.0 و بعد به نسخه 3.1 میرسند، ما هم امیدواریم که سایرین در ادامه این پروژه مشارکت کنند و فصل های جدیدی به آن اضافه کرده و یا فصلهای موجود را تغییر دهند. برای ممکن ساختن این هدف، نسخه الکترونیکی این کتاب را به طور رایگان و همینطور در گیتهاب⁴ هم ارائه می کنیم، بنابراین این کتاب می تواند با تلاش جمعی کامل شود.⁵

با تشکر از دوستانی که این کتاب را بازبینی کردند: اد فلتون⁶، ناتالی ایوانز-هریس⁷، فرانسواز کولت⁸ و کیسی فیسلر⁹.

⁴ GitHub

⁵ نسخه الکترونیک زبان اصلی کتاب روی سایت آمازون به صورت رایگان قابل دسترسی است.

⁶ Ed Felton

⁷ Natalie Evens-Harris

⁸ FranCois Chollet

⁹ Casey Lynn Fiesler

فصل 1 : انجام درست علم داده / علم داده درست / علم داده خوب ؟

بخش سخت یک دانشمند داده اخلاق گرا بودن فهم اخلاق نیست، بلکه پیوند بین ایده های اخلاقی و عمل و «انجام درست علم داده» است.

اخیرا بحث های خوب و سازنده ای در مورد اخلاق داده انجام شده است. از نظر ما این بحث ها مفید و ضروری هستند؛ اما مهمترین مساله ای نیستند که با آن مواجهیم. ما همین الان هم استانداردهای خوبی برای اخلاق داده داریم. کد اخلاقی ACM، که به سال 1993 بر می گردد و الان در حال به روزرسانی است، شفاف، دقیق و به طور شگفت آوری آینده نگرانه است. این کد اخلاقی، بعد از 25 سال، هنوز هم برای هر کسی که می خواهد در مورد اخلاق فکر کند، نقطه شروع مناسبی است. انجمن آماری آمریکا¹⁰ هم یک مجموعه راهنمای خوب برای کار با داده ها دارد. بنابراین لازم نیست از صفر شروع کنیم. ما معتقدیم بیشتر افراد می خواهند منصف باشند. دانشمندان داده و توسعه دهندگان نرم افزار قصد آسیب رساندن به افرادی که از محصولات آنها استفاده می کنند را ندارند. البته استثناهایی وجود دارد، که آنها را مجرم و هنرمندان فریبکاری می نامیم. تعریف "انصاف" در این مساله، کاری سخت و با در نظر گرفتن ابعاد متفاوت آن حتی غیرممکن است. با این حال نیازی نیست که این "انصاف" را از پیش دقیقاً تعریف کنیم و به علاوه نمی توانیم آن را به صورت یک بیانیه ساده اصول اخلاقی مطرح کنیم.

مسئله ای که در این کتاب با آن مواجهیم متفاوت از تعریف اصول اخلاقی است، مساله ما این است که چطور اصول اخلاقی را عملیاتی کنیم؟ ما درباره تعهد انتزاعی به منصف بودن حرف نمی زنیم. اصول اخلاقی اگر عملکرد ما را تغییر ندهند و هیچ اثری روی آنچه روزانه انجام می دهیم نداشته باشند بی فایده خواهند بود. این برای دانشمندان داده، چه آنهایی که به روش کلاسیک مشغول تحلیل داده هستند و چه آنهایی که تحلیل های پیشروی مبتنی بر هوش مصنوعی انجام می دهند، یک چالش بزرگ است. ما باید بدانیم که چطور سیستم های نرم افزاری بسازیم که انصاف را پیاده سازی کنند. منظور ما از «انجام درست علم داده» این است.

هر آیین نامه ای درباره اخلاق داده به شما خواهد گفت که نباید بدون رضایت آگاهانه سوژه، اطلاعات او را جمع آوری کنید. اما به شما نخواهد گفت که چطور این "رضایت آگاهانه" را کسب کنید. کسب رضایت آگاهانه وقتی دارید برای یک آزمایش روانشناختی با تعداد محدودی از افراد، شخصا مصاحبه می کنید ساده است. اما وقتی فردی در یک سایت خرید اینترنتی (آمازون) روی یک کالا کلیک می کند و بعد از آن تبلیغات مربوط به آن کالا در فضای لایتنهای تبلیغات شروع به تعقیب آن فرد می کنند، رضایت آگاهانه معنی متفاوتی دارد. آیا اگر شما بودید برای کسب اجازه آنها در جهت استفاده از کالاهای انتخابی شان در تبلیغات هدفمند یک پاپ آپ¹¹ باز میکنید؟ اگر همچین کاری بکنید چه تعداد مشتری را از دست خواهید داد؟ رضایت آگاهانه باز هم معنای متفاوتی خواهد داشت اگر از افراد بخواهید که برای عضویت در یک شبکه اجتماعی برای خود پروفایل بسازند و این داده در آینده ممکن است در اهداف آزمایشی استفاده بشود (ممکن هم هست از این داده استفاده ای نشود). آیا در این صورت فرم رضایتی را که حاوی اصطلاحات حقوقی غیرقابل درک است و عملاً می گوید "ما از داده شما استفاده خواهیم کرد اما نمی دانیم برای چه هدفی" به صورت پاپ آب برای کاربران باز می کنید؟ یا این موافقتنامه را به صورت یک انصراف نامه می نویسید و جایی پنهانش می کنید که کسی نتواند پیداش کند؟

¹⁰ The American Statistical Association

¹¹ pop-up

اینها نمونه ای از سوالاتی است که باید جواب بدهیم. و باید راهی پیدا کنیم تا بهترین راه حل ها و رویکردها را باهم به اشتراک بگذاریم. بعد از اصول اخلاقی باید به پیاده سازی اصول اخلاقی فکر کنیم. این کار ساده ای نیست و همه چیز را از طراحی تجربه کاربری تا مدیریت داده ها در بر میگیرد. تجربه کاربری را چطور طراحی کنیم که دغدغه هایمان در باره اخلاق و انصاف، یک اپلیکیشن¹² را غیرقابل استفاده نکند؟ همانقدر که نشان دادن یک پاپ آپ با هزاران کلمه از متن حقوقی غیرقابل درک ممکن است بد به نظر برسد، احتمال اینکه توضیحات دقیق و طولانی خسته کننده منجر به کسب رضایت و اجازه کاربر شود هم کم است.

چطور اطلاعات حساسی که بدست می آوریم را مدیریت می کنیم؟ گفتن اینکه اپلیکیشن ها نباید اطلاعات نژاد، جنسیت، معلولیت، یا سایر اطلاعات محافظت شده افراد را جمع آوری کنند راحت است. اما اگر این اطلاعات را جمع آوری نکنید، برای ارزیابی اینکه اپلیکیشن شما انصاف را در مورد اقلیتها رعایت می کند یا نه به مشکل برخورد خواهید خورد. به علاوه یادگیری ماشین نشان داده که حتی در صورت عدم ثبت اطلاعاتی مانند نژاد و اطلاعاتی از این قبیل، میتواند با استفاده از متغیرهای دیگر به خوبی آنها را تشخیص دهد. اپلیکیشن شما اولین اپلیکیشنی نخواهد بود که علی رغم نیت خوب توسعه دهندگانش منصفانه عمل نمیکند. آیا داده ای که برای تست منصفانه بودن سیستم تان لازم دارید را در یک پایگاه داده جداگانه با کنترلهای دسترسی جداگانه نگه می دارید؟

برای عملی کردن اصول اخلاقی باید فضا اخلاقی باشد. باید بتوانیم درباره اینکه اخلاق چیست، چه هزینه ای دارد و چه راه هایی برای پیاده سازی آن داریم گفتگو کنیم. ما به عنوان تکنولوژیست، معمولاً در کنفرانسها، پست های وبلاگهایمان و همچنین از طریق توسعه تکنولوژی های متن باز، بهترین راه حلها و رویکردهایمان را باهم به اشتراک میگذاریم. اما به ندرت در مورد مسائلی مثل اینکه چطور رضایت آگاهانه کاربران را به دست بیاوریم بحث می کنیم. این بحث چند بعد دارد که لازم است در مورد آنها فکر کنیم :

اول از همه نیاز به فرهنگ های سازمانی داریم که در آن، به بحث درباره انصاف، به بحث درباره استفاده درست از داده و صدماتی که با استفاده نادرست از داده ممکن است به بار بیاید اهمیت داده شود. و این به نوبه خود، یعنی باید از انتشار شتابزده محصولات قبل از فکر کردن به اینکه چطور استفاده خواهند شد اجتناب کنیم. ما نمی توانیم با بهانه اینکه "دوره دوره اینترنت است" نتایج و عواقب کارمان را نادیده بگیریم. ما پیامد نادیده گرفتن این نتایج در حیطه امنیت کامپیوتر را قبلاً تجربه کرده ایم: بسیاری از شرکتها که برای پیاده سازی شیوه های امنیتی و محافظتی، زمان لازم را اختصاص نداده اند، با لطمه ای که به شهرتشان خورده و ضرر مالی درگیر هستند. نگاهمان به مسائلی مثل انصاف، پاسخگویی و پیامدهای غیر منتظره باید مثل رویکردمان به امنیت باشد.

به خصوص لازم است درباره پیامدهای غیرمنتظره نحوه استفاده مان از داده ها فکر کنیم. البته پیش بینی تمام این عواقب غیرمنتظره غیرممکن است. ما انسان هستیم و توانایی مان برای پیش بینی آینده محدود است. اما بسیاری از عواقب غیرمنتظره به راحتی قابل پیش بینی بودند: مثلاً، قابلیت "سالی که گذشت" فیس بوک که مرگ ها و اتفاقات دردناک را به مردم یادآوری کرد. شعار "سریع حرکت کن و چیزها را بشکن" اگر به این فکر نکنیم که چه چیزهایی ممکن است شکسته شوند غیرقابل قبول است. و ما برای فکر کردن به این عواقب به فضای مناسبی احتیاج داریم. فضایی در زمانبندی پروژه، و همین طور که امکان اینکه به مدیران بگوییم لازم است در مورد یک محصول بازنگری کنند.

¹² Application

همچنین به فضایی نیاز داریم که وقتی یک اشکالی پیش می آید بتوانیم خط تولید را متوقف کنیم. این ایده به سیستم کانبان تویوتا¹³ برمیگردد: هر کارگر خط مونتاژ اگر متوجه اشکالی در کار می شد، می توانست آن خط را متوقف کند و آن خط مونتاژ تا زمانی که مشکل حل نمی شد، دوباره راه نمی افتاد. کارگرها ترسی از عواقبی که بخاطر متوقف کردن خط تولید ممکن بود از طرف مدیران در انتظار آنها باشد نداشتند. آنها مورد اعتماد بودند و انتظار می رفت مسئولانه رفتار کنند.

اگر بخواهیم مشابه این کار را در مورد ویژگی های محصولات نرم افزاری استفاده کنیم چه خواهد شد؟ اگر کسی در فیسبوک می توانست بگوید " صبر کنید ما در باره " سالی که گذشت " چند تا اعتراض دریافت کردیم " و توسعه آن را تا وقتی که کسی می توانست بررسی کند چه اتفاقی دارد می افتد متوقف کند؟

تصور کردن خشم مدیران در برخورد با چنین حرکتی راحت است. اما تصور یک "دکمه توقف" مشابه رویکرد تویوتا هم سخت نیست. فیسبوک به طور مداوم در حال استقرار¹⁴ محصولش است و آنها اغلب درباره اینکه چطور نیروهای جدید می توانند در روز اول کارشان تغییراتی در تولید محصول ایجاد کنند مانور میدهند. پس چرا اجازه ندهیم کارمندان توسعه برخی از ویژگی های محصول را متوقف کنند؟ چه ابزارهایی برای توقف فوری استقرار محصول وجود دارد؟ حتما چنین ابزارهایی وجود دارند، استقرار مداوم اگر نتوانیم آخرین تغییراتی که درست کار نکرده اند را به عقب برگردانیم معنا ندارد. بله فیس بوک یک شرکت بزرگ و پیچیده با یک محصول بزرگ و پیچیده است. همانطور که تویوتا یک شرکت بزرگ و پیچیده است و این روش برای آنها کار کرده است.

مساله ای که در پس همه این بحث ها قرار دارد، فرهنگ سازمانی است. محیط سازمان ممکن است نسبت به هر چیزی به جز سودآوری کوتاه مدت نگرش منفی داشته باشد. این ممکن است عواقب تصمیمات کوتاه نظرانه و دکتترین ضعیف اقتصادی باشد، به خصوص در ایالات متحده. و این ناگزیر ما را به سمت بزرگترین مساله راهنمایی می کند: چطور فرهنگ سازمان را تغییر بدهیم. سوزان اتلینگر¹⁵ پیشنهاد کرده است که در زمانی که بی اعتمادی و خشونت عمومی در حال بالا رفتن است، اخلاق یک سرمایه گذاری خوب است. مدیران سطح بالا به تازگی دارند این مشکل را می بینند، بنابراین تغییر فرهنگ سازمانی به سرعت رخ نخواهد داد.

کاربران مایلند با شرکت ها و سازمان های قابل اعتمادی که مطمئنند از آنها سوء استفاده نخواهند کرد تعامل کنند. کاربران می خواهند با شرکتهایی در تعامل باشند که با آنها و داده های آنها مسئولانه رفتار کنند و نه فقط برای بیشینه کردن سود ممکن برای جذب مشتری. اینها شرکتهایی خواهند بود که در سازمان خود برای اخلاق فضایی ایجاد می کنند.

ما، دانشمندان داده، مهندسان داده، توسعه دهندگان هوش مصنوعی و یادگیری ماشین و سایر متخصصانی که با داده کار می کنند، باید این تغییر را درخواست کنیم. ما نمی توانیم آن را به افرادی اختصاص دهیم که صرفا کارشان در حوزه اخلاق است. نمی توانیم از مدیریت انتظار داشته باشیم که چند نفر اخلاقگرا استخدام کنند و به تیم ما اضافه کنند. ما باید بر اساس ارزشهای اخلاقی زندگی کنیم، نه اینکه تنها در مورد آنها صحبت کنیم. باید در مورد عواقب کارهایمان به دقت فکر کنیم. ما باید فضای اخلاق را در سازمان مان ایجاد کنیم. تغییرات فرهنگی ممکن ست زمانبر باشند، اما در نهایت اتفاق خواهند افتاد اگر ما خودمان تغییر نماییم. این معنای انجام درست علم داده است.

¹³ Toyota's Kanban

¹⁴ Deployment

¹⁵ Susan Etlinger

فصل 2 : سوگندها و چک لیست ها

"سوگندها؟ ما به هیچ کدام از سوگندهای نفرت انگیز نیازی نداریم!" (با عرض معذرت از هامفری بوگارت¹⁶ در گنج های سیرا مادره¹⁷)

سال گذشته، بحثهای زیادی درباره اخلاق داده انجام شد، که انگیزه بسیاری از آنها، نگرانی ها در رابطه با اخبار جعلی¹⁸، تبلیغات هدفمند، جهت گیری¹⁹ الگوریتمی و تاثیراتی که محصولات مبتنی بر داده بر روی افراد و جامعه دارند، بود. نگرانی ها درباره اخلاق داده تازه نیست؛ IEEE²⁰، ACM²¹ و انجمن آماری آمریکا²²، همگی قوانین اخلاقی مرتبط با داده ها دارند. اما شدت بحث های اخیر نشان می دهد که اتفاق مهمی در حال وقوع است: علم داده دارد به بلوغ می رسد و متوجه مسئولیت های خود می شود. یک جهان بهتر فقط بخاطر اینکه ما از داده ها استفاده می کنیم نخواهد آمد. داده هم یک وجه تاریک دارد. با الهام از سوگند قدیمی بقراط برای پزشکان، بحث درباره اخلاق داده معمولا به سمت سوگندهای داده ای منحرف می شود. ما ضمن اینکه از توان و فکری که برای این سوگندها گذاشته می شود قدردانی می کنیم، درباره ارزش آن مردد هستیم. چرا که سوگندها چند مساله دارند :

+ سوگندها یکبار اتفاق می افتند. شما یک بار سوگند می خورید و تمام (اگر اصلا سوگند بخورید!). و هیچ دلیلی وجود ندارد که هر لحظه به یاد آن باشید. شما هر روز سوگندها را برای خودتان تکرار نمی کنید. یا به طور منظم بررسی نمی کنید که آیا بر اساس ایده آل ها زندگی می کنید یا نه.

+ سوگندها مجموعه ای از اصول کلی و گسترده هستند. بحث ها درباره سوگند بقراطی با عبارت " اول، آسیبی وارد نکن" شروع می شود، کلماتی که در واقع در سوگند وجود ندارند. اما "ابتدا، آسیبی وارد نکن" به چه معناست؟ برای قرنهاي متمادی دکترها به جز آسیب رساندن کار چندانی برای بیمار انجام نمی دادند. (افراد زیادی فقط به این خاطر که پزشکان اعتقادی به شستن دست هایشان نداشتند جان خود را از دست دادند). پزشکان نمی دانستند که در حال آسیب وارد کردن به بیمارانشان هستند. سوگند ایده خوبی است اما در اجرا کافی نیست. و در علم داده نیز (همچون پزشکی) همه چیز به نحوه عملکرد مربوط می شود.

+ سوگندها حتی ممکن است به عنوان پوششی برای افراد یا سازمانهایی که کار غیراخلاقی می کنند عمل کند. این نحوه فکر کردن که "کار ما نمی تواند غیراخلاقی باشد چون ما این سوگند را یاد کرده ایم" راحت است. گفتن اینکه "بدجنس نباش"²³ کافی نیست. شما باید واقعا بدجنس نباشید.

+ در سوگندنامه ها هیچ تلاشی برای ایجاد ارتباط بین اصول و عمل انجام نمیشود. گفتن اینکه محققان باید از افراد رضایت آگاهانه بگیرند" یک بحث است و کسب رضایت آگاهانه در ابعاد اینترنت بحث دیگری است. و آموزش چپستی رضایت آگاهانه به کاربران یک بحث کاملا متفاوت است.

منظور ما این نیست که اصولی که در این سوگندها آمده است، مهم نیستند، مشکل آنها این است که ما را به جایی که می خواهیم نمی رسانند. سوگندها ایده ما از اینکه چه چیزهایی خوب و منصفانه هستند را به روشهای پیاده سازی آنها وصل نمی

¹⁶ Humphrey bogart

¹⁷ Treasure of the Sierra Madre

¹⁸ Fake news

¹⁹ Bias

²⁰ ACM Code of Ethics and Professional Conduct

²¹ IEEE Organizational Ethics

²² American Statistical Association's Ethical Guidelines for Statistical Practice

²³ "Don't be evil" : عبارتی است که در آیین نامه رفتاری گوگل آمده است و قبلا هم به عنوان شعار این شرکت استفاده می شد.

کنند. ما می توانیم ساعتها در باره انصاف و بی طرفی صحبت کنیم بدون اینکه بدانیم چطور باید منصف و بی طرف باشیم. در این حالت سوگندها حتی خطرناک هم می شوند. چرا که شما را متقاعد می کنند که شما آدم خوبی هستید و در حال انجام کار درست هستید. در حالی که در واقع اینطور نیست.

سوگندها برای شروع بحث خوب هستند و در سال گذشته بحث هایی زیادی ایجاد کرده اند. این بحث ها در آگاه ساختن مردم درباره مسائلی مثل بی طرفی الگوریتمیک²⁴ بسیار مفید بوده اند. این بحث ها به توسعه دهندگان نرم افزار و دانشمندان داده در درک اینکه کار بی تاثیر نیست و تاثیر واقعی (چه خوب یا چه بد) روی مردم دارد کمک کرده است. بحث های شدیدی در خصوص اینکه استقلال برای دانشمندان داده به چه معناست و چه قوانینی بیشتر از چند سال دوام خواهند داشت وجود دارد، اما باید قدم بعدی را برداریم و ایده ها را به عمل تبدیل نماییم. اما چگونه؟

در سال ۲۰۰۹، اتول گواند²⁵ کتاب کوتاهی با موضوع اینکه چگونه از اشتباهات بزرگ اجتناب کنیم به نام بیانیه چک لیستها (انتشارات مک میلان)²⁶ نوشت. او چیزهای زیادی از راهکارهایش به عنوان یک جراح نوشت. در یک بیمارستان همه میدانند که باید چکار کنند. همه میدانند که شما موظفید که قبل از عمل ضد عفونی شوید. همه میدانند که شما نباید پای دیگر را به اشتباه قطع نمایید. همه میدانند که هنگام بخیه زدن شما نباید اسفنج و بقیه تجهیزات را درون بدن بیمار جا گذاشته باشید.

با اینحال باز هم خطا و اشتباه رخ می دهد، علی الخصوص وقتی شما در شرایط پر از استرس قرار دارید. جراح به اشتباه پای دیگر را عمل میکند. اسفنج داخل بدن بیمار جا میماند و چیزهای از این قبیل. گواند عاقبت فهمید که با ساختن چک لیستی از مواردی که باید حواستان به آنها باشد، به سادگی می توان از تمام این اشتباهات اجتناب کرد. هرچند، پزشکی هستند که ایده ساختن چک لیست را به طرز توهین آمیزی ساده میبینند. و این گروه همانهایی هستند که پیوسته اشتباه میکنند.

چک لیستها، برخلاف سوگندنامه، قوانین را به عمل تبدیل میکنند. هر کسی میداند باید قبل از جراحی ضد عفونی انجام شود. این یک قانون است. اما اگر مجبور شوید که بعد از انجام هر کار، گزینه مربوط به آن را در یک فرم تیک بزنید، احتمال فراموش کردن این کار بسیار کم است. این یک راهکار است. علاوه بر این چک لیست فقط یکبار خوانده نمی شوند. چک لیستها سوگندنامه ای نیست که یکبار در مراسم تشریفاتی خوانده شوند. بلکه راهکاری است که در هر رویه ای با آنها سر و کار خواهیم داشت.

یک چک لیست برای یک دانشمند داده یا یادگیری ماشین به چه شکل خواهد بود؟ چارچوب اخلاق داده دولت انگلستان و کتاب کار اخلاق داده²⁷ یک راهکار است. آنها ۷ تا قانون را مشخص کرده اند و برای هر کدام جزییاتی را ارائه داده اند. این کتاب کار تعدادی پرسشهای بدون جواب قطعی مطرح کرده است که پذیرش شما در برابر این قوانین را محک بزند. نقد ما این است که فرآیند پیشنهادی آنها سربار زیادی را تحمیل میکند. هرچند کسی که به طور کامل فرآیند آنها را دنبال کند، قطعاً به طور دقیق راجع به موضوعات اخلاقی فکر خواهد کرد، اما در عمل، درخواست از توسعه دهندگان برای پر کردن کتاب کار با جوابهای طولانی به ۴۶ تا سؤال، میتواند تضمین کند که تفکر اخلاقی اتفاق نخواهد افتاد.

ما معتقدیم که چک لیستها حول و حوش سؤالات ساده ای مانند «آیا این کار انجام شده است» ساخته می شوند و دلیل کارا بودن آنها همین سادگی آنهاست. درواقع آنها فضایی برای مانور ندارند. یا شما تشخیص داده اید که یک پروژه چگونه میتواند

²⁴ algorithmic fairness

²⁵ Atul Gawande

²⁶ The Checklist Manifesto (Macmillan)

²⁷ The UK Government's Data Ethics Framework and Data Ethics Workbook

مورد سوءاستفاده قرار گیرد یا خیر. یا مکانیزمی برای گرفتن رضایتنامه ایجاد نموده‌اید یا خیر! درست است که تضمینی بر بی‌نقص بودن این راهکار نیست: تحلیل شما ممکن است نامناسب باشد و مکانیزم جلب رضایت شما ایراد داشته باشد، اما حداقل «ی‌توانید ادعا کنید که آن را انجام داده‌اید».

در خصوص استفاده و تغییر این چک لیست در پروژه هایتان آزاد هستید. این چک لیست بسیاری از اصولی که در سوگندنامه های داده‌ای در موردش بحث می‌شود را پوشش میدهد. هنگام شروع یک پروژه، چک لیست را مرور نمایید تا توسعه دهندگان بدانند چه چیزهایی لازم است و با مجموعه جدیدی از نیازمندیها در دقایق آخر، مواجه نشوند. وقتی نرم‌افزار آماده شد، از آن استفاده نمایید. آن را در دست بگیرید و دقیق همه گزینه ها را قبل از ارائه محصول به عموم چک کنید.

نمونه‌ای از چک لیست برای کسانی که روی پروژه های داده‌ای کار میکنند به شرح زیر است:

- ❑ آیا لیستی از روشهایی که این تکنولوژی می‌تواند مورد تهاجم یا سوءاستفاده قرار بگیرد را فهرست کرده ایم.
- ❑ آیا منصفانه و جامع بودن داده‌های آموزشی را مورد آزمون قرار داده ایم؟
- ❑ آیا سرچشمه های احتمالی جهتدار بودن داده²⁸ را کاملاً بررسی نموده ایم؟
- ❑ آیا تیم ما تنوع نظرات، پیش‌زمینه ها و انواع تفکرات را بازتاب داده است؟
- ❑ به چه نوع رضایتی از کاربران نیاز داریم تا از داده استفاده نماییم؟
- ❑ آیا مکانیزمی برای کسب رضایت از کاربر فراهم کرده ایم؟
- ❑ آیا شفاف شرح داده‌ایم که از کاربر در چه زمینه هایی رضایت گرفته میشود؟
- ❑ در صورتی که نتیجه کار، آسیبی به کسی وارد کند آیا مکانیزمی برای خسارت‌دهی در نظر گرفته ایم؟
- ❑ آیا میتوانیم در صورت رفتار نادرست نرم افزار، آن را در زمان استفاده متوقف نماییم؟
- ❑ آیا رعایت انصاف را در گروههای مختلف کاربری ارزیابی کرده ایم؟
- ❑ آیا نرخهای خطای متفاوت به ازای گروههای کاربری متفاوت را بررسی نموده ایم؟
- ❑ آیا در خصوص انقضای مدل تست و بررسی نموده ایم تا مطمئن باشیم نرم‌افزار ما در طول زمان منصف باقی میماند؟

❑ آیا برنامه‌ای برای پشتیبانی و ایمن نگاه داشتن داده کاربر داریم؟

سوگند نامه و آیین نامه هر یک ارزش خود را دارند. ارزش سوگند به متعهد شدن نیست، بلکه به فرآیندی است که از طریق آن در توسعه سوگند پیش می‌روید. افرادی که با داده‌ها کار میکنند بحثهایی دارند که در دهه گذشته هرگز وجود نداشته اند. اما این بحثها باعث انجام شدن کارهای سخت نمیشوند، و ما باید به سمت کارهای سخت برویم. ما نمیخواهیم در مورد چگونگی استفاده اخلاقی از داده‌ها صحبت کنیم بلکه میخواهیم از داده‌ها به صورت اخلاقی استفاده کنیم. ریاکارانه است که در مورد اخلاق صحبت کنیم، اما در راستای آن کاری انجام ندهیم. ما میخواهیم قوانین را به عمل تبدیل نماییم. و اینجاست که چک لیستها به ما کمک خواهند نمود.

²⁸ Bias in data

فصل سوم پنج اصل توسعه محصولات داده ای

برای ایجاد یک سرویس یا محصول داده خوب، به چه چیزهایی نیاز است؟ منظور فقط مفید بودن یا از لحاظ تجاری ماندگار بودن محصول نیست، بلکه استفاده از داده ها به صورت اخلاقی و مسئولیت پذیرانه است.

ما غالباً در خصوص تکنولوژی یک محصول یا تجربه کاربری²⁹ آن صحبت میکنیم ولی به ندرت راجع به چگونگی ساختن یک محصول داده که مسئولیت پذیرانه، کاربر را در مرکز توجهاتش قرار دهد صحبت میکنیم. نیاز به این نوع محصولات بسیار زیاد است. شنیدن اخبار مربوط به اینکه مردم به محصول داده‌ای که از آن استفاده میکنند – یا درواقع محصول از آن‌ها استفاده میکند- «اعتماد ندارند» متداول است. هرچند فیسبوک بیشترین پوشش خبری را به خود اختصاص داده است، اما عدم اعتماد به یک پلتفرم خاص محدود نیست. کمبود اعتماد اکنون همه گیر شده است و شامل تقریباً هر شرکت تجارت آنلاین، خرده فروشان بزرگ، گردآورندگان و دلالان داده و در صنعت و دولت می شود.

کاربران اعتماد خود را از دست داده اند. به دلیل اینکه بارها احساس کرده‌اند مورد سوءاستفاده قرار گرفته اند؛ تبلیغات مخرب، مشاهده محتوای دروغین و نادرست، و مشاهده فرهنگ «اول کارت را بکن، بعد تا میتوانی عذرخواهی کن» در بسیاری از شرکت‌های آنلاین بارها این حس را به آن‌ها القاء نموده است. حتی طبیعی است کاربران این حس را به دلیل بسیاری سوءاستفاده هایی که حتی از آن خبر ندارند داشته باشند. چرا درخواست خسارت آنها از بیمه تایید نشد؟ چرا با درخواست وامشان موافقت نمیشود؟ آیا این تصمیمات توسط سیستمی گرفته شده است که با داده‌های جهندار شده آموزش دیده است؟ شعار میگوید "سریع حرکت کن و چیزها را بشکن"، اما اگر جامعه آسیب ببیند چطور؟

جمع آوری داده تجارت بزرگی است. داده ارزشمند است: «سوخت جدید» نامی است که اقتصاددانان برای آن در نظر گرفته اند. ما مدتی است که آن را میدانیم. اما جامعه داده را با این فرض در اختیار قرار میدهد که ما (همان جامعه) قرار است از آن سود ببریم. همچنین فرض ما این است که داده به صورت مسئولانه جمع آوری و نگهداری می‌شود و کسانی که داده را فراهم نموده اند آسیبی نخواهند دید. اساساً این مدلی از اعتماد است. اما وقتی اعتماد شکسته شد چگونه میتوان آن را دوباره ایجاد نمود؟ فایده‌ای ندارد که وانمود کنید قابل اعتماد هستید، در حالیکه در عمل خلاف آن را ثابت نموده اید. تنها راه برای بازگرداندن اعتماد، این است که واقعاً قابل اعتماد شوید و بازیابی این اعتماد وقتی آن را از دست دادید طول خواهد کشید.

هیچ راه ساده‌ای برای بازیابی اعتماد کاربر وجود ندارد، اما ما میخواهیم یک قانون طلایی برای داده را به عنوان نقطه شروع پیشنهاد نماییم: «با داده دیگران همانطوری رفتار کنید که از سایرین انتظار دارید با داده شما رفتار کنند.» هر چند پیاده‌سازی این قانون طلایی در فرآیند واقعی تحقیق و توسعه، چالش برانگیز است-همانطور که رسیدن از یک سوگند کوتاه و ناپایدار و تعهد به عمل واقعی دشوار است.

این جمله که با داده دیگران مانند داده خودتان عمل کنید به چه معناست؟ چند تا از دانشمندان داده واقعا در مورد این که چگونه از داده آن‌ها ممکن است استفاده و یا سوءاستفاده شود فکر کرده اند؟ و وقتی که فهمیدید که میخواهید با داده هایتان(و داده‌های دیگران) چطور رفتار شود، چگونه آن را پیاده‌سازی خواهید نمود؟ قانون طلایی به تنهایی کافی نیست. ما نیاز به دستورالعمل‌هایی داریم که بحث‌هایی را به تیم توسعه اپلیکیشن، کاربران اپلیکیشن و افراد دیگری که با استفاده از داده جمع آوری شده در معرض آسیب قرار دارند، تحمیل نماید.

²⁹ User Experience

پنج تا دستور العمل ساختاری به ما کمک میکند که در مورد ساختن محصولات داده‌ای تفکر نماییم. ما نام آن‌ها را پنج اصل توسعه محصولات داده‌ای³⁰ می‌گذاریم. رضایت، شفافیت، سازگاری، کنترل (و شفافیت) و پیامدها (و آسیبها). این‌ها چارچوبی برای پیاده‌سازی قانون طلایی داده هستند. بیایید به هر کدام از آن‌ها نگاهی بیندازیم.

رضایت³¹

بدون توافق بر سر اینکه چه داده‌ای از کاربر جمع‌آوری می‌شود و این داده چگونه استفاده خواهد شد نمی‌توان بین افرادی که داده را فراهم می‌کنند و آنهایی که از داده استفاده می‌کنند اعتماد ایجاد کرد. این توافق با کسب رضایت درباره جمع‌آوری و استفاده از داده شروع می‌شود. متأسفانه موافقت نامه‌ها بین کاربران یک سرویس (افرادی که داده آنها جمع‌آوری می‌شود) و آن سرویس (که از داده جمع‌آوری شده به طرق مختلف استفاده می‌کند) صفر و یکی و فاقد **وضوح** هستند. در تجارت، وقتی دو طرف یک قرارداد درباره آن مذاکره می‌کنند، معمولاً قبل از اینکه به نتیجه نهایی برسند با توجه به خط قرمزهای طرفین، چند دور مذاکره صورت می‌گیرد. اما وقتی یک کاربر می‌خواهد بر سر یک قرار داد با یک سرویس داده موافقت کند، باید با تمام بندهای قرارداد موافقت کند در غیر این صورت به آن سرویس دسترسی نخواهد داشت. بندهای این قرارداد غیرقابل مذاکره هستند.

مثلاً وقتی در یک بیمارستان پذیرش می‌شوید، باید فرمهایی پر کنید که به آنها اجازه میدهد از داده‌های شما استفاده کنند و راهی نیست که بگویید می‌خواهید از داده شما برای یک سری از اهداف استفاده شود و در سایر اهداف نه. وقتی برای دریافت کارت وفاداری (کارت عضویت) داروخانه محل تان امضا می‌دهید، توافق می‌کنید که آنها می‌توانند از داده شما هر استفاده‌ای بکنند. این استفاده‌ها حتماً شامل تبلیغات هدفمند (معمولاً تحت عنوان پیشنهادات ویژه) خواهد بود، اما می‌تواند شامل فروش اطلاعات شما (با یا بدون بینام‌سازی³² داده‌ها) به شرکتهای دیگر هم باشد. چه اتفاقی برای داده‌های شما خواهد افتاد اگر یک شرکت، شرکت دیگر را بخرد و از اطلاعات شما برای اهدافی که اصلاً مورد انتظار شما نبوده است استفاده کند؟

داده‌ها مرتباً بدون کسب رضایت، جمع‌آوری، استفاده و فروخته می‌شوند. از جمله توسط سازمان‌هایی مثل اکسیوم³³، اکسیفکس³⁴، اکسپرین³⁵ و ترنس یونیون³⁶ که با هدف ارزیابی ریسک مالی اقدام به گردآوری داده می‌کنند. با اینحال بسیاری از برندهای متداول هم بدون کسب رضایت اقدام به جمع‌آوری داده می‌کنند. در اروپا، گوگل از دوربین‌هایی که روی اتومبیل‌ها نصب شده بودند، برای توسعه محصولات مسیریابی خود، داده جمع‌آوری کرد. AT&T و کامکست³⁷ هر دو از گیرنده‌های دیجیتال کابلی برای جمع‌آوری اطلاعات کاربرانشان استفاده کردند و سامسونگ³⁸ صدای ضبط شده توسط دستگاه‌های تلویزیونی که دستور صوتی دریافت می‌کردند را جمع‌آوری کرد. نمونه‌های خیلی خیلی بیشتری از گردآوری

³⁰ در نسخه اصلی از این اصول به نام پنج C یاد می‌شود. چرا که عنوان هر پنج اصل با حرف C شروع می‌شود: consent, clarity, consistency, control (and transparency), and consequences (and harm)

³¹ Comsent

³² Anonymization: بی‌نام‌سازی یا ناشناس‌سازی، فرآیند رمزگذاری یا حذف اطلاعاتی در مجموعه داده است که اشخاص را قابل شناسایی شخصی می‌کند. پس از انجام این فرآیند افرادی که اطلاعات آن در مجموعه داده‌ها آمده است به صورت ناشناس باقی می‌مانند.

³³ Acxiom

³⁴ Equifax

³⁵ Experian

³⁶ Transunion

³⁷ Comcast

³⁸ Samsung

داده بدون کسب رضایت وجود دارند. در هر مرحله از ساختن یک محصول مبتنی بر داده، پرسیدن این سوال که آیا رضایت لازم و مناسب درباره داده از کاربر گرفته شده است ضروری است.

صداقت و روراستی³⁹

روراستی رابطه نزدیکی با رضایت دارد. شما نمی توانید به درستی درباره چیزی رضایت بدهید، مگر اینکه به طور آشکار به شما گفته بشود درباره چه چیزی رضایت می دهید. کاربران باید به وضوح بدانند چه داده ای را فراهم می کنند، با این داده چه کاری انجام خواهد شد و هرگونه پیامد ناشی از نحوه به کار بردن داده های خود را بدانند. اغلب مواقع توضیح اینکه چه داده ای گردآوری یا فروخته می شود در اسناد حقوقی طولانی که به ندرت به دقت خوانده می شوند (اگر اصلاً خوانده شوند) دفن می شوند. مثلاً خوانندگان هوشیار توافقنامه کاربر ایونت برای⁴⁰ اخیراً متوجه شدند که ثبت یک رویداد در این شرکت حق ارسال یک تیم فیلمبرداری و حق انحصاری ویدیوی ضبط شده درباره آن رویداد را به این شرکت می داد و تنها راه انصراف از دادن چنین حقی این بود که در این باره درخواستی کتبی برای این شرکت ارسال کنید. وقتی مردم فهمیدند که قبول کردن این توافقنامه چه تبعاتی میتواند داشته باشد، واکنش هایشان بسیار سریع بود و ایونت برای این بند را حذف کرد. کاربران فیسبوک که بازی "این زندگی دیجیتالی شماست"⁴¹ متعلق به کمبریج آنالیتیکا⁴² را بازی کرده اند ممکن است متوجه شده باشند که با این کار اطلاعات خود را به این شرکت تقدیم می کردند، چرا که آنها به سوالاتی جواب می دادند و این جواب ها حتماً در جایی ذخیره می شد. اما آیا این افراد می دانستند از داده آنها چه استفاده ای ممکن است بشود؟ یا می دانستند که در اصل آنها داشتند اجازه دسترسی به اطلاعات دوستانشان در فیس بوک را هم به این شرکت می دادند؟ این نکته در لایه های عمیق تنظیمات محرمانگی فیسبوک دفن شده است.

کاربران حتی با اینکه از عمومی بودن فرومهایی که در آنها داده میگذارند آگاهند، غالباً نمی دانند از داده های آنها چه استفاده هایی خواهد شد. بیشتر کاربران توئیتر⁴³ می دانند که توئییت های عمومی آنها در دسترس عموم قرار می گیرد. اما اغلب آنها نمی دانند که توئییت های آنها می تواند گرد آوری و در کارهای تحقیقاتی استفاده شود، یا حتی اینکه مجموعه توئییت ها به فروش می رسد. این به این معنا نیست که چنین استفاده ای از داده های کاربران غیر اخلاقی است، بلکه همانطور که کیسی فیذر⁴⁴ اشاره می کند، باید در کنار تلاش برای کسب رضایت کاربران، آنها را نسبت به موضوعی که درباره آن رضایت می دهند آگاه کرد و این به معنای صداقت و روراستی است.

تقریباً در تمام سرویس ها، به ندرت پیش می آید که حتی یک توضیح ساده درباره اینکه آن سرویس با داده جمع آوری شده چه خواهد کرد و این استفاده چه پیامدهایی ممکن است داشته باشد، به کاربران داده شود. متأسفانه معمولاً فرایند کسب رضایت در جهت مبهم کردن جزئیات و عواقب آنچه کاربر نسبت به آن رضایت می دهد استفاده می شود و یکبار که داده ها از دست رفتند دیگر نمی توان آنها را بازگرداند. شما نمی توانید داده های که در اختیار سرویس دهنده قرار داده اید را پس بگیرید. حتی اگر سازمانی بخواهد داده ها را از بین ببرد، اثبات اینکه این کار واقعاً انجام شده باشد بسیار دشوار است.

³⁹ Clarity، در این بخش از آنجایی که منظور نویسندگان بیان روشن و واضح زمینه هایی است که داده جمع آوری شده از کاربران در آنها استفاده خواهد شد.

⁴⁰ Eventbrite: ایونت برای عنوان یک شرکت و وبگاه است که تمرکز اصلی آن روی مدیریت رویدادها و ارائه خدمات فروش برخط بلیت برای رویدادهای گوناگون است. مراجعه کنندگان به وبگاه ایونت برایت می توانند رویدادهای نزدیک به محل خود را بیابند

⁴¹ This is your Your Digital Life

⁴² Cambridge Analytica

⁴³ Twitter

⁴⁴ Casey Fiesler

البته چند استثنای قابل توجه وجود دارد: افرادی مثل جان ویلبانکس⁴⁵ روی مدلهایی که به کاربران در درک بهتر پیامدهای انتخابشان کمک می کنند کار می کنند. کارهای ویلبانکس به کاربران کمک می کند که اتفاقاتی را که با ارائه داده های حساس پزشکی و سلامتی به یک سرویس می افتد را درک کنند.

ثبات و اعتماد⁴⁶

ثبات در طول زمان لازمه اعتماد است. ما نمی توانیم به افرادی که غیر قابل پیش بینی هستند اعتماد کنیم. چنین افرادی ممکن است بهترین نیت ها را داشته باشند، اما درست زمانی که انتظارش را ندارید، آن نیت ها را نادیده بگیرند. یا حتی چنین نیت هایی را به روشی عجیب و غیرقابل پیش بینی تفسیر کنند. و اعتماد وقتی از بین رفت، دوباره به دست آوردن آن ممکن است زمان زیادی ببرد. بازگرداندن اعتماد نیاز به یک دوره طولانی از رفتار سازگار دارد.

ثبات، و در نتیجه اعتماد ممکن است مستقیم یا به صورت ضمنی از بین بروند. سازمانی که داده های کاربران را در معرض آسیب قرار می دهد، ممکن است این کار را به عمد یا ناخواسته انجام داده باشد. در سالهای اخیر، شاهد حوادث امنیتی زیادی بوده ایم که در آنها داده کاربران دزدیده شده: انتم⁴⁷، تارگت⁴⁸، یاهو⁴⁹، بیمارستانهای محلی، داده های دولتی، دلایلی داده ای مثل اکسپیرین⁵⁰ و این لیست هر روز طولانی تر می شود. شکست در محافظت از داده های کاربران منجر به از بین رفتن اعتماد می شود- و محافظت از داده ها اگر در طول زمان ثبات نداشته باشیم بی معناست.

ما همچنین شاهد سرخوردگی، ترس و تعجب کاربرانی بوده ایم که به درستی متوجه نشده اند که با چه چیزی توافق کرده اند. وقتی کمبریج آنالیتیکا⁵¹ از داده های فیسبوک کاربران استفاده کرد و با استفاده از آسیب پذیری های کاربران تبلیغات ویژه به آنها نمایش داد، فیسبوک ابتدا اعلام کرد که این یک نشت اطلاعات نبوده است. و با اینکه به لحاظ فنی، حق با فیسبوک بود که داده ها دزدیده نشده اند اما ادراک عمومی با این نظر متفاوت بود. این شاید تعرض فیسبوک به حقوق کاربران نبود اما قطعا تخطی از اعتماد عمومی بود. فیسبوک در اجرای موافقت نامه ای که با مشتریان خود داشت با ثبات نبود. وقتی اخبار منتشر شد، فیسبوک غیرقابل پیش بینی شد، چرا که بیشتر کاربران نمی دانستند چه اتفاق خواهد افتاد. آنها توافقنامه کاربر و تنظیمات پیچیده حریم شخصی در فیسبوک را درک نمی کردند و نمی دانستند فیسبوک این تنظیمات را چگونه تفسیر خواهد کرد.

کنترل و شفافیت⁵²

وقتی داده ای را به یک سرویس می دهید باید بدانیم چه اتفاقی برای این داده شما خواهد افتاد. آیا می توانید نحوه استفاده این سرویس از داده خود را کنترل کنید؟ مثلا فیسبوک دیدگاه سیاسی، مذهبی، و گرایش جنسی شما را می پرسد (البته الزامی نیست). اگر نظرتان درباره داده ای که به فیس بوک داده اید عوض شد چه؟ اگر تصمیم گرفتید دیدگاه سیاسی خودتان را پنهان نگه دارید، آیا می دانید فیسبوک اطلاعاتی که در این زمینه از شما دارد را حذف می کند یا نه؟ و یا همچنان از این اطلاعات در انتخاب آگهی ها استفاده می کند؟

⁴⁵ John Wilbanks

⁴⁶ Consistency and Trust

⁴⁷ Anthem

⁴⁸ Target

⁴⁹ Yahoo

⁵⁰ Experian

⁵¹ Cambridge Analytica

⁵² Control and Transparency

بیشتر مواقع کاربران کنترل موثری بر چگونگی استفاده از داده هایشان ندارند. به آنها انتخاب های همه یا هیچ داده می شود یا مجموعه پیچیده ای از انتخاب ها در اختیارشان قرار داده میشود که کنترل دسترسی به داده ها را برای آنها پیچیده و گیج کننده می کند. حذف برخی یا تمام داده ای که توسط کاربر در اختیار سرویسها قرار گرفته است اغلب غیر ممکن است. یک تغییر مهم در حقوق حریم خصوصی داده این است که به سمت دادن کنترل بیشتر کاربران بر داده هایشان پیش می رود. برای مثال مقررات حفاظت از اطلاعات عمومی اروپا (GDPR)⁵³، ارائه داده به کاربر در صورت درخواست او و همچنین حذف داده های کاربر در صورت تمایلش را الزامی می کند.

پیامدها⁵⁴

محصولات داده ای با هدف ایجاد ارزش برای یک کاربر یا سیستم ایجاد می شوند. با پیچیده تر شدن این محصولات و گسترش تاثیر اجتماعی آنها این سوال اساسی پیش می آید که آیا داده ای که توسط آنها جمع آوری شده است می تواند به افراد و گروه ها صدمه ای وارد کند یا نه؟ ما درباره عواقب پیش بینی نشده و موارد ناشناخته استفاده از داده ها و ترکیب مجموعه داده های متفاوت مدام می شنویم. ریسک ها را هرگز نمی توان حذف کرد. اما بسیاری از این پیامدهای پیش بینی نشده و موارد ناشناخته می توانست پیش بینی شوند اگر تنها تلاش کرده بودیم. دلیل ناشناخته بودن این موارد، این است که ما نمی خواهیم آنها را بشناسیم.

با توجه به استفاده های بالقوه ای که از داده می توان کرد، قانون و سیاست های ویژه ای برای حفاظت از گروه های خاص وضع شده است: مثلاً قانون حریم خصوصی آنلاین کودکان (COPA)⁵⁵ که از کودکان و داده های آنها محافظت می کند. همینطور قوانینی برای حفاظت از مجموعه داده های حساس خاص وجود دارد: مثلاً قانون عدم تبعیض ژنتیکی (GINA)⁵⁶ که در 2008 در واکنش به افزایش نگرانی ها در مورد استفاده از تست ژنتیکی علیه یک فرد یا خانواده او تصویب شد. متأسفانه سیاست گذاری، پا به پای پیشرفت های تکنولوژی پیش نمی رود و هیچ یک از این دو قانون به روز رسانی نشده اند. با توجه به رشد به کارگیری تکنولوژی توسط جامعه، دولت اوباما به این نتیجه رسید که سرعت تنظیم مقررات نمی تواند با سرعت تکنولوژی همگام شود. در نتیجه، سمت هایی با عنوان مسئول ارشد تکنولوژی و دانشمند ارشد داده معرفی کرد. دولت اوباما بیش از 40 نقش مسئول ارشد داده و دانشمند داده در دولت فدرال ایجاد کرد. در نتیجه ضمن پرورش و حمایت از نوآوری، در تنظیم مقررات، از بررسی آسیبهای احتمالی، زود به زود و به طور منظم اطمینان حاصل می شود.

حتی رویکردهای بشردوستانه هم ممکن است پیامدهای ناخواسته و مضر داشته باشند. وقتی AOL در سال 2006، اطلاعات جستجو های انجام شده توسط این موتور جستجو را به صورت بینام، برای استفاده پژوهشگران منتشر کرد، ثابت شد که می توان با انجام فرآیند معکوس بینام سازی روی این اطلاعات برخی از کاربران را شناسایی کرد. در سال 2018، استراوا⁵⁷ داده هایش را باز کرد تا کاربران بتوانند مکان های جدیدی برای دویدن یا دوچرخه سواری پیدا کنند. استراوا نمی دانست که اعضای ارتش آمریکا هم از گجت های پوشیدنی که GPS فعال داشتند استفاده می کردند و فعالیت آنها پایگاه های نظامی و مسیرهای بازرسی آنها در عراق و افغانستان را در معرض دید بقیه قرار می داد. انتشار این اطلاعات بعد از اینکه این محصول به عموم ارائه شد مشخص شد و افراد پس از بررسی اطلاعات شروع به صحبت درباره نگرانی هایشان کردند.

⁵³ General Data Protection Regulation

⁵⁴ Consequences

⁵⁵ the Children's Online Privacy Protection Act

⁵⁶ Genetic Information Nondiscrimination Act

⁵⁷ Strava

گرچه AOL و استراوا یک زنجیره از اتفاقات پیش بینی نشده را با انتشار این داده ها کلید زدند، باید بدانیم که این داده ها حتی اگر به طور عمومی منتشر نمی شدند هم می توانستند خطرناک باشند. جمع آوری داده های ظاهرا بی ضرر و ترکیب آن با مجموعه داده های دیگر ممکن است پیامدهای جدی داشته باشد. ترکیب مجموعه داده های مختلف معمولاً منجر به نتایجی می شود که بسیار قوی تر و خطرناکتر از هر یک از مجموعه داده ها به تنهایی است. مثلاً داده های مسیر دویدن و دوچرخه سواری اگر با اطلاعات قفل های هوشمند ترکیب می شد، می توانست به سارقان بگوید یک آپارتمان یا خانه کی و برای چه مدت خالی است. این داده ممکن بود توسط یک هکر دزدیده می شد و این شرکت هرگز متوجه این آسیب نشود.

انتقاد از استراوا و AOL و گفتن اینکه استراوا نباید این محصول را منتشر می کرد یا AOL نباید اطلاعات جستجو را منتشر می کرد ساده است ولی آن روی سکه را نادیده می گیرد. در هر دو مورد دانشمندان داده با نیت خیرخواهانه به دنبال کمک به دیگران بودند. مشکل اینجاست که عواقب و ریسک های احتمالی کارشان را کاملاً بررسی نکرده بودند.

میتوان داده را برای کارهای پژوهشی به اشتراک گذاشت و همزمان از عدم داشتن عواقب ناخواسته مطمئن بود. مثلاً خدمات درآمد داخلی آمریکا⁵⁸، با همکاری پژوهشگران، مجموعه داده مشابهی را با یک روش کاملاً کنترل شده برای کمک به درک نابرابری اقتصادی باز کرد. هیچ تأثیر منفی یا پیامد سیاست گذاری اشتباهی رخ نداد. همینطور اداره حمل و نقل، داده ای درباره تلفات ترافیکی منتشر کرد. بیوبانک انگلستان⁵⁹ (یکی از بزرگترین مجموعه داده های ژنوم) یک روش بسیار پیچیده برای در اختیار قرار دادن داده در سطوح مختلف دارد. شرکت های دیگری هم هستند که موفق شده اند داده هایشان را برای منافع عمومی باز کنند. از جمله پروژه گراف اقتصادی لینکدین⁶⁰ و نمایش دهنده n-gram گوگل بوک.

مجموعه داده های زیادی هستند که می توانند منافع عظیمی ایجاد کند و با این حال در سرورها مانده اند. داده های پزشکی که در چندین موسسه مختلف پخش شده است سرعت پژوهش را کم می کند. داده بدست آمده از به اشتراک گذاشتن تاکسی و همین طور داده شرکت های نقشه برداری GPS می تواند به ازدحام ترافیکی و ایمنی حمل و نقل کمک کند. اما در اختیار قرار دادن داده ها برای اهداف پژوهشی باید با برنامه صورت پذیرد.

پیاده سازی پنج اصل توسعه محصولات داده ای

استفاده از داده ها می تواند زندگی عادی ما را به روشهای مختلفی به یک زندگی شگفت انگیز تبدیل کند. مثلاً سیستم پیشنهاد فیلم ابزار خیلی خوبی است، یا اگر می توانستیم داده های پزشکی بیماران کل دنیا را در جایی به صورت متمرکز داشته باشیم میتوانستیم پیشرفت چشمگیری در درمان بیماریهایی مانند سرطان داشته باشیم. اما مادامیکه از پیاده سازی کارآمد پنج اصل توسعه محصولات داد های اطمینان نداشته باشیم، بهبود سیستم پیشنهاد فیلم یا پیشرفت در درمان سرطان عملی نخواهد شد. اگر با داده های دیگران مثل داده های خودمان محتاطانه برخورد نکنیم بهبودی در هیچ کدام از این سیستم ها نخواهیم داشت. در دهه گذشته صنعت نرم افزار تلاشهای قابل توجهی برای بهبود تجربه کاربری (UX) انجام داده است. بخش عمده این تلاشها در راستای توسعه رویکردهای کاربر محور برای ساختن محصولات و سرویس هایی بوده است که وابسته به داده

⁵⁸ the US International Revenue Service (IRS)

⁵⁹ The UK Biobank

⁶⁰ LinkedIn's Economic Graph project

هایی است که از کاربران جمع آوری میشود. نتیجه همه این تلاشها این بوده است که استفاده از نرم افزارها به طور کلی ساده تر و لذت بخش تر شده است. اما متأسفانه، تلاش این تیمها خواسته یا ناخواسته محدود به جلب رضایت لحظه ای کاربران یا امکان تحقق اهداف کوتاه مدت بوده است، همانطور که تمرکز «هک رشد» روی جذب افراد برای ثبت نام در خدمت رسانها از طریق روشهای همه گیر است. تیمهای کمی را دیده ایم که در توسعه تجربه کاربری تلاش کنند بین تجربه های لحظه ای کاربر و ارزشهای بلند مدت توازن برقرار کنند.

به طور خلاصه، تیمهای توسعه و تولید، تأثیرات پنج اصل معرفی شده برای توسعه محصولات داده ای را در نظر نگرفته اند. برای مثال یک برنامه چگونه باید به کاربران اطلاع دهد که از داده هایشان چه استفاده ای خواهد شد و چگونه رضایتشان را جلب کند. این بخش از تجربه کاربری نباید نادیده گرفته شود و به این معنی هم نیست که رضایت گرفتن از کاربر را آسان و «نه» گفتن را دشوار کنیم. این مسئله هم بخشی هم تجربه کاربری است. کاربران باید بدانند در مورد چه چیزی رضایت میدهند و این رضایت چه تأثیری خواهد داشت، در غیر اینصورت کار طراح تجربه کاربری ناقص خواهد بود.

البته مسئولیت این پنج اصل فقط بر عهده طراحان تجربه کاربری نیست و کل تیم در این قضیه مسئولند. دانشمندان داده باید با پرسیدن سؤالاتی «چه می شود .. اگر» که همه این پنج اصل را دربرگیرد با این مسئله روبرو شوند. مدیران محصول، رهبران کسب و کار، مدیران فروش و بازاریابی و مدیران اجرایی نیز همین مسئولیت را بر عهده دارند.

پنج اصل مذکور باید جزیی از فرهنگ هر سازمانی باشد و محصول و طراحی آن باید به طور منظم برای بررسی این 5 اصل بازبینی شوند. سازمانها باید پیش از معرفی یک محصول به عموم، ایجاد چک لیست مناسب را مد نظر داشته باشند. غالباً ما تا حد زیادی محصولات داده ای را به صورت MVP (محصولات قابل توسعه مینیمال) میبینیم. (MVP: الگوهایی که تست می کنند آیا یک محصول برای کاربران ارزشمند است یا خیر). هرچند این روش برای تست و توسعه ایده های جدید سازنده است، ولی حتی MVP هم باید پنج اصل توسعه محصولات داده ای را رعایت کند. این قضیه برای محصولات خوش ساخت نیز صادق است. روشهای جدیدی ممکن است توسعه داده شوند که به صدمات غیر قابل پیش بینی منجر شوند. به طور خلاصه، بحث در خصوص به عهده گرفتن مسئولیت محصولاتی که ساخته شده اند می باشد. این پنج اصل مکانیزمی برای بهبود مکالمه با کاربر به منظور اطمینان دادن در خصوص آسیب رسان نبودن محصول است.

فصل چهارم

روز تسویه حساب داده

زندگی ما با داده ها عجین شده است. از سیستمهایی که به ما پیشنهاد می دهند در شبکه های اجتماعی با چه کسی "دوست" شویم یا چه کسی را "دنبال" کنیم تا وسایل نقلیه خودمختار داده محور. اما تجربه سالهای گذشته نشان داده است که محصولات و تکنولوژی هایی که ساخته ایم به عنوان سلاحی علیه خودمان استفاده شده است. هرچند استفاده از داده ها از جهات بسیاری برای ما سودمند بوده است، اما تضادهایی هم بین حریم خصوصی، مصلحت جمعی و سود سازمانی ایجاد کرده است.

کتاب سلاح های کشتار ریاضیاتی⁶¹ کتی اونیل⁶² (انتشارات برادوی بوکز⁶³) و کتاب خودکارسازی نابرابری⁶⁴ از ویرجینیا اوبانک⁶⁵ (انتشارات مک میلان⁶⁶) روشهای زیادی که در آنها از داده ها برای آسیب رساندن به قسمت وسیعی از جامعه استفاده شده است را مستند کرده اند.

علم داده، یادگیری ماشین، هوش مصنوعی و تکنولوژیهای وابسته به آنها اکنون به روز تسویه حساب رسیده اند. زمان آن رسیده است که مسئولیت آنچه خلق کرده ایم را برعهده بگیریم. اما به عهده گرفتن مسئولیت ایجاد، نگهداری، مدیریت داده و تکنولوژیها و سرویسها به چه معناست؟ مسئولیت، ناگزیر با انگیزه های پیچیده ای که منجر به تولید یک محصول می شوند در هم تنیده است. این انگیزه ها در جبهه و مرکز بحثهای مرتبط با نقش شبکه های اجتماعی در انتخابات 2016 آمریکا، جذب نیروی تروریست و حملات آنلاین، قرار دارد. دیگر کاملاً روشن شده است که انگیزه سازمانهایی مالک محصولات مبتنی داده هستند یا آنها را تولید می کنند، در راستای منفعت کاربران این محصولات نبوده است. این مساله، برای کاربران اینترنت مساله جدیدی نیست. بقیه علوم هم روز تسویه حساب خود را گذرانده اند. برای مثال در پزشکی، سواستفاده هایی که به طور گسترده به آنها پرداخته شد عبارتند از آزمایشات سفیلیس تاسکیگی⁶⁷، مورد هنریتا لاکس⁶⁸ که سلول هایش (بدون اجازه و بدون پرداخت غرامت) برای تحقیق روی سرطان مورد استفاده قرار گرفتند و آزمایشات انسانی که در جنگ جهانی دوم بوسیله نازیها انجام شد. جامعه فیزیک باید با پیامدهای ساختن بمب اتم دست به گریبان می شدند. زیست شناسان و شیمی دانان باید درباره استفاده از تحقیقاتشان برای ساختن بمبهای زیستی و شیمیایی توضیح می دادند. بقیه رشته های مهندسی هم تأثیر کارهای غیر اصولی خود بر زندگی مردم را فهمیده اند، نمیتوان خراب شدن پلها را نادیده گرفت. در نتیجه همه اینها، گروههای متشکل از افراد خبره تشکیل داده شده اند که آیین نامه هایی را آماده و تصویب نمایند و فرایندهای نظارتی دولتی، استانداردها و مجازاتی برای کارهایی که برای جامعه خطرناک است تعیین نموده اند.

آموزش اخلاق و امنیت

در بسیاری از رشته ها، اخلاق جزء ضروری آموزش حرفه ای است. اما در علم کامپیوتر، علم داده، هوش مصنوعی و سایر رشته های وابسته چنین نیست. دروس اخلاق در مدارس زیادی تدریس میشوند، اما موضوعاتی که در این کلاسها تدریس میشوند به پروژه های موجود و تمرینات درسی ارتباطی ندارد. دانشجویان ممکن است به صورت تئوری، قوانین اخلاقی را یاد بگیرند اما یاد نمی گیرند که در عمل چطور این اصول را در پروژه هایشان بکار ببرند. در نتیجه برای چالشهای دنیای واقعی آماده نیستند. آنها برای فکر کردن به موارد اخلاقی و نحوه تأثیر آن بر انتخابشان آموزش ندیده اند. آنها نمیدانند که چگونه درباره پروژه ها یا تکنولوژیهایی که ممکن است منجر به ضررهایی در دنیای واقعی بشوند بحث کنند.

امنیت نرم افزار و اخلاق، غالباً به هم وابسته اند و روش فعلی ما برای آموزش امنیت، مثال خوبی از آنچه نباید در آموزش اخلاق انجام بدهیم اختیارمان قرار می دهد. امنیت عموماً به صورت درسی اختیاری و جدا از سایر کلاسهای مرتبط به توسعه نرم افزار تدریس میشود. برای مثال در کلاس درس پایگاه داده هیچگاه در خصوص حملات sql-injection حرفی زده نمی شود. علیرغم نیاز به آموزش آنها در کلاسهای پایگاه داده، این نوع حمله در کلاسهای امنیت بررسی میشوند. وقتی

⁶¹ Weapons of Math Destruction

⁶² Cathy O'Neill

⁶³ Broadway Books

⁶⁴ Automating Inequality

⁶⁵ Virginia Eubanks

⁶⁶ Macmillian

⁶⁷ Tuskegee : شهری در ایالت آلابامای امریکا

⁶⁸ Herietta Lacks

دانشجویی پروژه ای را در درس پایگاه داده انجام میدهد، میزان آسیب پذیری آن در مقابل حملات مخرب در نمره اش تأثیری ندارد. در سیستم نمره دهی خودکار حتی آسیب پذیری آن پروژه در برابر این حملات بررسی نمی شود. علاوه بر این در درس پایگاه داده درباره طراحی معماری برای کم کردن تخریب وارده به سیستم در صورتی که یک مهاجم بتواند به سیستم نفوذ کند هم صحبتی نمی شود. (برای مثال ذخیره عناصر داده ای مانند نام و شماره های امنیتی در پایگاه داده های مجزا)

هرچند آموزش امنیت در قالب دروس انتخابی بهتر از عدم آموزش آن است، ولی بهترین روش برای تربیت برنامه نویسهایی که واقعا امنیت را بفهمند، این است که علاوه بر تدریس آن در دروس اختیاری، رعایت آن در تمرینها و نمره دهی دروس اصلی دوره هم تأثیر داده شود. با این روش تضمین میشود که همه بتوانند مسایل پایه ای امنیتی شناسایی و آنها را حل نمایند. دروس انتخابی هم میتوانند عمق بیشتری به این موضوع داده و آن را با محدوده وسیعی از موضوعات- از امنیت فیزیکی تا رمزنگاری- ترکیب نمود. امنیت به عنوان موضوعی که پس از توسعه محصول به آن فکر شود موثر خواهد بود بنابراین چرا ما انتظار داریم آموزش آن به این صورت موثر باشد؟ صنعتی وجود ندارد که وقفه های امنیتی در آن به دزدیدن داده منجر نشده باشد و در نتیجه زندگی میلیونها نفر تحت تأثیر قرار نگرفته باشد. شیوه های امنیتی ضعیف منجر به آسیب پذیری جدی در لوازم مصرفی از قفلهای هوشمند تا لامپهای حبابی هوشمند شده است.

اخلاق نیز با مساله مشابهی روبرو است. اخلاق داده در بسیاری از کالج ها و دانشگاهها تدریس میشود اما از بقیه دروس تحصیلی جدا است. هرچند این درسها، به دانشجویان کمک میکند که به طور جدی به این مباحث بیندیشند، اما نمیتواند در حوزه مسایلی مانند گرفتن رضایت آگاهانه در محدوده یک برنامه واقعی کمکی کند. گزارش کاخ سفید با عنوان «آماده شدن برای آینده هوش مصنوعی»⁶⁹ بر نیاز به آموزش در هر دو بخش اخلاق و امنیت تأکید می کند:

"آموزش اخلاق برای متخصصان هوش مصنوعی و دانشجویان، بخش ضروری از راه حل این مساله است. به طور ایده آل، هر دانشجویی که در یکی از رشته های علوم کامپیوتر، هوش مصنوعی و علم داده تحصیل میکند، باید با دروس و مباحث مرتبط با اخلاق و امنیت روبرو شود. البته، اخلاق به تنهایی کافی نیست. گرچه اخلاق کمک میکند که متخصصان مسئولیت خود را در قبال همه ذینفعان بدانند، اما در عمل برای بهبود، آموزش اخلاق باید به ابزارها و روشهای فنی افزوده شود تا بوسیله انجام کارهای فنی لازم از نتایج غیرقابل قبول جلوگیری شود."

اخلاق و امنیت باید به جای بودن در لیست دروس انتخابی یا دروس مجزا، در برنامه های درسی اصلی قرار گیرند. باید در کلیه درسهای کالج ها، دانشگاهها، آموزشهای آنلاین و دوره های برنامه نویسی لحاظ شوند. باید به جای انتزاعی بودن، در کنار «ابزارها و روشهای فنی جهت قدرت بخشیدن به عملیات» قرار گیرد و این آموزش نباید پس از فارغ التحصیلی متوقف شود. کارفرماها باید به طور منظم انجمنهایی را میزبانی نموده و دروس جدیدی را پیشنهاد دهند که افراد را در جریان آخرین چالشها و چشم اندازها قرار دهند.

توسعه اصول راهنما

مساله اصول اخلاقی این است که فراموش کردن و نادیده گرفتن آنها وقتی عجله دارید کار ساده ایست: وقتی پروژه را باید در یک زمانبندی فشرده و شاید حتی غیرواقع بینانه به پایان برسانید. وقتی عقربه های ساعت لحظه به لحظه به زمان تحویل نزدیک می شوند، فراموش کردن همه چیزهایی که در کلاس آموزش دیده اید ساده است، حتی اگر آن کلاس اخلاق و راه حل های مسایل دنیای واقعی را به هم مرتبط می کرده است.

⁶⁹ Preparing For The future Of Artificial Intelligence , Executive Office of The president, National Science and Technology Council Committee on Technology, October 2016

چک لیست ها، روش ثابت شده ای برای حل این مساله هستند. چک لیست ها همانطور که اتول گواند⁷⁰ در مانیفست چک لیست⁷¹ (انتشارات متروپولیتن بوکز⁷²) گفته است، کم کم جزیی از آیین و روش کار می شوند. چک لیست یک مجموعه کوچک از سوال هاست که در ابتدای پروژه و سپس در هر مرحله ای که به راه اندازی پروژه نزدیک می شوید دوباره از خود می پرسید و تنها در صورتی مرحله بعدی را شروع می کنید که جواب تان به تمام سوالات چک لیست مثبت باشد. ثابت شده است که چک لیست ها، خطاها را در عمل جراحی کاهش می دهند، خلبانهای خطوط هوایی هم خیلی از آنها استفاده می کنند، به خصوص در شرایط اضطراری. به همین طریق چک لیستها می توانند به دانشمندان و متخصصان داده هم کمک کنند که مسایل اخلاقی را فراموش نکنند، حتی وقتی برای تحویل سریعتر محصول تحت فشار باشند.

در **فصل دوم**، چک لیستی برای توسعه دهندگان نرم افزارهای داده محور پیشنهاد دادیم. در استفاده و یا تغییر این چک لیست برای آنکه با وضعیت شما سازگار شود و سپس استفاده از آن در پروژه تان راحت باشید. چک لیست ما تمام مسائلی که باید در نظر بگیرید را دربر نمی گیرد. این چک لیست حتی تمام برنامههایی که در حال توسعه هستند را در بر نمی گیرد چه رسد به برنامه هایی که در آینده توسعه داده خواهند شد. با اینحال اگر موردی به آن اضافه کردید، سعی کنید که آن را کوتاه نگه دارید. چرا که کوتاه بودن موارد یک چک لیست است که آن را کارا می کند.

گروه عدالت، پاسخگویی و شفافیت در یادگیری ماشین⁷³ (FAT/ML) روش مشابهی را پیشنهاد میکند. "اصول الگوریتمهای قابل حسابرسی و بیانیه تاثیر اجتماعی برای الگوریتم ها"⁷⁴ که توسط این گروه ارائه شده است توصیه می کند تاثیر اجتماعی یک پروژه حداقل 3 بار در طول عمر پروژه ارزیابی شود: در زمان طراحی، پیش از راه اندازی و پس از راه اندازی. کار بر اساس بیانیه تاثیر اجتماعی مستلزم آن است که توسعه دهندگان نرم افزار درباره عواقب اخلاقی پروژه خوب فکر کنند و هر مشکلی در این باره که شناسایی کردند را حل کنند. به روش مشابه، اصول جمعی شیوه های اخلاقی داده⁷⁵ که حاصل کنفرانس داده برای تبادل خوب⁷⁶ (D4GX) است مجموعه ای از ارزشها و اصول را که در بحث های این جامعه جمع آوری شده اند را ارائه می دهد. این اصول شروع خوبی برای هر گروهی است که می خواهند چک لیست خودشان را ایجاد کنند. در کنار اینها، کتی اونیل برای تحقق عدالت در الگوریتم های یادگیری ماشین، بررسی و نظارت بر آنها را پیشنهاد کرده است.

تزریق اخلاق به فرهنگ داده محور

مسئولیت فردی کافی نیست. اخلاق باید بخشی از فرهنگ سازمانی باشد. سازمانهای زیادی را دیده ایم که ارزش توسعه یک فرهنگ داده محور را به رسمیت می شناسند. باید اطمینان حاصل کنیم که اخلاق و امنیت هم بخشی از آن فرهنگ بشوند. امنیت، رفته رفته بخشی از فرهنگ سازمانی می شود چرا که پیامدهای حرفه ای، مالی، حقوقی و اعتباری قربانی [حملات امنیتی] شدن بسیار بزرگتر از آن هستند که نادیده گرفته شوند. سازمانها در حال آزمایش برنامه های جایزه باگ (باگ باونتی⁷⁷)، به اشتراک گذاری تهدیدها و همکاری با سازمانهای دولتی هستند. تیم های امنیت دیگر فقط ناظر سازمان نیستند، بلکه مسئول پیشگیری از آسیبهای جدی بر شهرت و وضعیت مالی سازمان هستند.

⁷⁰ Atul Gawande

⁷¹ The Checklist Manifesto

⁷² Metropolitan Books

⁷³ www.fatml.org

⁷⁴ Principles for Accountable Algorithms and a social Impact Statement for Algorithms

⁷⁵ The Community Principles on Ethics Data Practices

⁷⁶ Data for Good Exchange

⁷⁷ bug-bounty programs

ادغام اخلاق در فرهنگ سازمانی، چالش برانگیزتر است. یکی از اعضای تیم ممکن است به یک روش معترض باشد، اما پذیرفته نشدن اعتراض این فرد و کنار گذاشته شدنش از تیم ساده است. و اگر در یک سازمان، از تفکر اخلاقی حمایت نشود، احتمالاً نتیجه اعتراض همین خواهد بود. تفکر اخلاقی با یا بدون حمایت و پشتیبانی سازمان مهم است، اما وقتی اقدام اخلاقی در سازمان ارزش باشد، احتمال تغییر بیشتر خواهد بود. اینجا چند ایده درباره تزریق اخلاق به فرهنگ سازمانی را ذکر می کنیم:

باید به افراد این قدرت داده شود که یک فرایند را قبل از اینکه باعث آسیبی بشود، متوقف کنند. توپوتا و دلیو. ادوارد دمیگ⁷⁸ در استفاده از ریسمان اندون⁷⁹ برای بهبود کیفیت و کارایی پیشرو بودند. هر فردی که متوجه مشکلی می شد می توانست ریسمان را بکشد و خط تولید را متوقف کند. سپس مدیران ارشد و کارگران خط تولید مساله را به بحث می گذاشتند، مشکل را حل می کردند و تولید را از سر می گرفتند. هر یک از اعضای تیم داده هم باید بتواند هر زمان که مشکلی مشاهده کردند، با کشیدن یک ریسمان مجازی، توسعه را متوقف کند. محصول یا ویژگی های جدید تا زمانی که راه حلی برای آن مشکل پیدا شود، متوقف می ماند. به این ترتیب، یک فرآیند تکرار شونده می تواند توسعه یابد که مانع از نادیده گرفتن مشکلات می شود.

هر کسی باید بدون ترس از تلافی کردن (تنبيه)، مشکلات را با هدف اصلاح آنها برجسته کند. باید فرایندی برای اعلام مشکلات وجود داشته باشد تا اعضای تیم که حس می کنند نظرشان نادیده گرفته شده است بتوانند اظهار نظر کنند. وزارت امور خارجه ایالت متحده امریکا، یک کانال مخالفت⁸⁰ دارد که هر دیپلماتی از طریق آن می تواند مطمئن باشد که وزیر امور خارجه نگرانی ها و نظرات او را خواهد شنید. در مراقبت های بهداشتی، قانون وجود روشی برای بیان و اعلام مسائل حقوقی و اخلاقی را الزامی کرده است. در برنامه های مراقبت های بهداشتی در امریکا، یک افسر انطباق⁸¹ وجود دارد که مستقیماً به هیئت مدیره گزارش می دهد.

سازمانهای داده محور هم به مدل مشابهی نیاز دارند که به افراد اجازه دهد بدون نگرانی از انتقام، مسائل و مشکلات را مطرح کنند. این فرآیند مطرح کردن مشکلات را می شود به چند روش مختلف پیاده سازی کرد. مثلاً شرکتها می توانند با سازمانی مثل بنیاد مرزهای الکترونیکی⁸² (EFF) همکاری کنند و برنامه ای برای دریافت و رسیدگی به گزارشهای افشاگری ایجاد کنند که مساله گزارش شده را تا زمانی که معیارهای خاصی نقض نشود رسانه ای نخواهد کرد. روش مشابهی تحت نظارت یک اداره جدید یا یکی از آژانس های موجود، می تواند در پیش گرفته شود. (مثلاً یک اداره حفاظت از اطلاعات مصرف کننده⁸³)

در فرآیند استخدام باید یک چالش اخلاقی هم گنجانده شود.

⁷⁸ W. Edwards Deming

⁷⁹ andon cord : یک ابزار مدیریتی برای اصلاح رسانی درباره مشاهده هرگونه مشکل در محصول یا فرآیند تولید آن است.

⁸⁰ dissent channel

⁸¹ Compliance officer :A **compliance officer** is an individual who ensures that a company complies with its outside regulatory and legal requirements as well as internal policies and bylaws

⁸² Electronic Frontier Foundation

⁸³ Consumer Data Protection Agency

موقع استخدام، شرکت ها اغلب "تناسب فرهنگی" کارمند جدید را ارزیابی می کنند. مصاحبه کننده ها معمولاً سؤالی طرح می کنند تا مطمئن شوند کارمند جدید می تواند به خوبی با سایر اعضای تیم کار کند. با اینحال، به ندرت در مصاحبه ها سؤالی در باره ارزشهای اخلاقی از مصاحبه شونده پرسیده می شود. به جای پرسیدن سؤالاتی که جواب آنها بله/خیر است، دریافتیم این که بهتر است سؤالاتی مطرح کنیم که کمک می کند بفهمیم مصاحبه شونده درباره انتخابهای مربوط به اخلاق و امنیت چه طرز فکری دارد. این سؤالی است که ما از آن در مصاحبه استفاده می کنیم:

فرض کنید یک مجموعه بزرگ از اطلاعات جمعیت شناسی⁸⁴ (آمار) در اختیار داریم. می خواهیم افراد را ارزیابی کنیم و قرار نیست که از اطلاعاتی نژادی افراد به عنوان ورودی مدل استفاده کنیم. با اینحال شما یک متغیر واسطه برای نژاد پیدا می کنید و با استفاده از چند متغیر دیگر می توانید نژاد افراد را محاسبه کنید. چه کار خواهید کرد؟ آیا از آن استفاده خواهید کرد؟

چنین سؤالی یک بحث در باره نحوه استفاده از متغیرهای واسطه شروع خواهد کرد: این کار چه تاثیری روی افرادی که از این محصول استفاده می کنند خواهد داشت؟ آیا توصیه می کنیم یا تصمیم می گیریم این سرویس را ارائه بدهیم یا نه؟ آیا یک نیاز قانونی را پیاده سازی می کنیم؟ آیا یک راهنما در باره سازگاری تهیه می کنیم؟ بحث درباره این سوال و پاسخ های احتمالی، ارزشهای مصاحبه شونده را آشکار خواهد کرد.

در بررسی محصول، باید در باره تاثیرات محصول سوال شود.

بیانیه های پیامدهای زیست محیطی، تاثیرات پروژه های عمرانی را بر عموم پیش بینی می کنند. ما قبلاً از پیشنهاد بیانیه پیامدهای اجتماعی توسط FAT/ML به عنوان مثالی از کار مشابهی که برای داده ها می توان کرد، یاد کردیم. در علوم اجتماعی و صنعت زیست پزشکی، هیئت های بررسی نظارت⁸⁵ (IRBs)، پیامدهای آزمایشات را قبل از اجرای آنها ارزیابی می کنند.

با اینکه هر دو بیانیه تاثیرات زیست محیطی و IRBs مشکلات محصولات داده ای را ارائه می کنند، تیم های داده باید تاثیر انتخاب هایی که می کنند را بررسی کنند. این تیم ها باید به پیامدهای اعمال شان قبل از انتشار محصول فکر کنند. ما معتقدیم که استفاده از یک چک لیست بهترین روش برای حصول اطمینان از نتایج خوب است.

تیم ها باید در اندیشه ها، تجربیات، نژادها و پس زمینه ها تنوع داشته باشند

خیلی درباره محصولات که به لحاظ فرهنگی ناهنجار هستند و یا آشکارا نژادپرست هستند می شنویم. یک مثال بارز، سیستم کنترل پاسپورت اتوماتیکی است که تا وقتی یک تصویر دیجیتالی مناسب از افراد نگیرد، اجازه ادامه مسیر به آنها نمی دهد. در این سیستم افرادی که نژاد آسیایی داشتند، گزارش دادند که این سیستم دائماً از آنها می خواست که چشمان خود را باز کنند در حالی که چشمان آنها باز بود. مثال های خجالت آور زیادی شبیه به این ثبت شده اند و بسیاری از آنها را می توان ناشی از کمبود داده یا کمبود بینش درباره تنوع جمعیتی که تحت تاثیر آن قرار خواهند گرفت دانست.

با اینکه راه حل کلی ای برای این مسائل مربوط به فرهنگ نداریم، تنوع و گنجاندن افراد متنوع در تیم کمک بزرگی خواهد بود. افراد تیم باید جزئی از جمعیتی باشند که تحت تاثیر محصول قرار خواهند گرفت. در این صورت آنها مشکلات را قبل از اینکه هر کس دیگری ببیند، خواهند دید. نظارت خارجی هم می تواند کمک کند مشکلاتی که تیم آنها را نمی بیند شناسایی

⁸⁴ demographic data

⁸⁵ Institutional Review Boards

شوند. چرا که معمولا وقتی عمیقا درگیر پروژه ای هستید، مشکلاتی که از دید افراد خارج از پروژه بدیهی هستند را به سختی می بینید.

سازمان ها باید اصول خود را شفاف کنند

اصل "بدجنس نباش"⁸⁶ گوگل همیشه نمونه ای از یک اصل بامزه و در عین حال مبهم بوده است اما بیانیه اخیر آنها "هوش مصنوعی در گوگل: اصول م⁸⁷" واضح تر است. به همین ترتیب، استارتاپ شناسایی چهره کایروس⁸⁸ اعلام کرده است که با شرکتهای مجری قانون، همکاری نخواهد داشت. مدیر عامل کایروس می نویسد که : استفاده از تشخیص چهره تجاری در اجرای قانون یا هر یک از دستگاه های نظارتی دولت اشتباه است."

با اینحال باید بدانیم که حمایت از اصول اخلاقی شرکت ها عواقبی دارد. برای ایجاد اصول هوش مصنوعی، اعتراضات گسترده داخلی و استعفای تعدادی از توسعه دهندگان در اعتراض به قراردادهای دفاعی گوگل لازم بود. کایروس هم به احتمال زیاد پول زیادی را از دست میدهد. در کنار اینها، درک این نکته هم مهم است که سازمانها اغلب برای پنهان کردن پروژه های غیراخلاقی شان، مدام اصول اخلاقی شان را مطرح می کنند.

در چند سال اخیر در باره استارتاپ های نرم افزاری که با "محصول پذیرفتنی کمینه"⁸⁹ شروع می کنند و سپس به شعار "سریع حرکت کن و چیزها را بشکن"⁹⁰ فیس بوك می چسبند زیاد شنیده ایم. آیا این روش با راه حلی که ما ارائه دادیم ناسازگار است؟ انتخاب بین حرکت سریع و خراب کردن یا شکستن چیزها انتخاب اشتباهی است. حرکت سریع به معنای خراب کردن چیزها نیست و می توان سریع و مسئولانه حرکت کرد.

متدولوژی چابک/lean که اغلب در بسیاری از استارتاپ ها استفاده می شود راه مناسبی برای آشکار کردن مسائل اخلاقی قبل از اینکه تبدیل به مشکل بشوند است. توسعه دهندگان با یک محصول بسیار ساده (" ساده ترین چیزی که کار کند" بر اساس اصل اولیه ...) شروع می کنند، آن را به کاربر نمایش می دهند، بازخورد کاربر را دریافت می کنند، نسخه بعدی محصول را توسعه می دهند و این مراحل را تکرار می کنند. این فرایند به هر تعدادی که لازم باشد تکرار می شود تا زمانی که محصول بدست آمده راضی کننده باشد. اگر یک گروه متنوع از کاربران، محصول را ارزیابی کنند، چرخه توسعه محصول مشکلات سیستماتیک مربوط به پیش داور⁹¹ و ناهنجاری اجتماعی را احتمال زیاد آشکار خواهد کرد. کلید اصلی، تست و ارزیابی محصول توسط یک گروه واقعا متنوع از کاربران است نه گروهی که تنها بازتابی از کاربران مورد انتظار و یا توسعه دهندگان است.

آیین نامه

⁸⁶ Don't be evil

⁸⁷ Artificial Intelligence at Google: our principles

⁸⁸ Kairos

⁸⁹ Minimum Viable Product

⁹⁰ Move fast and break things

⁹¹ bias

در برخی صنایع، استانداردهای اخلاقی بوسیله قوانین و آیین نامه‌ها تحمیل می‌شود. «آیین نامه نورنبرگ»⁹² در پاسخ به جنایات نازیها توسعه داده شد و تمرکز آن بر روی رضایت فردی برای شرکت در یک آزمایش است. پس از اینکه «آزمایشات سیفلیس در تاسکیگی» رسانه ای شد، این آیین نامه در لایحه تحقیقات ملی 1947⁹³ و بیانیه 1975 هلسینکی⁹⁴ در قانون گنجائیده شد. این فشار برای آیین نامه کردن اصول اخلاقی باعث تاسیس «کمیته بازنگری سازمانی»⁹⁵ (IRB) شد و از طریق قانون عمومی⁹⁶ به طور گسترده ای در آمریکا پذیرفته شد.

در سایر صنایع، سایر سازمانهای تنظیم مقررات، استانداردهای سازمانی را اجرا می کنند. از جمله این سازمانها «کمیسیون تجارت فدرال» یا FTC⁹⁷ است که بر تجارت نظارت میکند، «کمیسیون تنظیم مقررات هسته ای» یا NRC⁹⁸ است که بر نیروگاههای هسته ای نظارت میکند، «سازمان مدیریت غذا و دارو» یا FDA⁹⁹ است که بر سلامتی دارویی نظارت میکند و اخیراً «دفتر حمایت مالی از مشتری» یا CFPB¹⁰⁰ است که از طرف مشتری وظیفه نظارت بر بانکدار و وام دهنده را بر عهده دارد.

آیین نامه حمایت از داده‌های عمومی (GDPR¹⁰¹) در اتحادیه اروپا یک رویکرد تهاجمی برای قانونمند کردن استفاده از داده‌ها و ایجاد یک خط مشی داده یکسان در پیش گرفته است. در ژوئن ۲۰۱۸، کالیفرنیا علیرغم بی میلی بسیاری از شرکتهای آنلاین یک قانون حریم دیجیتال مشابه با GDPR تصویب نمود. یک چالش توسعه چارچوب خط مشی این است که فرآیند توسعه آن تقریباً همیشه از سرعت نوآوری در تکنولوژی عقب می ماند و به اندازه کافی برای اینکه با آن هماهنگ باشد چابک نیست. تا زمانی که یک خط مشی مشخص و پذیرفته شود تقریباً همیشه از تکنولوژی عقب مانده است. اما برای سیاستگذاران ممکن نیست که کارشان را به اندازه کافی سریع تکرار کنند تا با جدیدترین تکنولوژی سازگار باشند. مشکل بعدی این است که اعضای سیاستگذار به اندازه کافی پیش زمینه فنی قوی ندارند. این میتواند خوب باشد چون تکنولوژیستها به راحتی تحت تاثیر نیازهای تکنولوژی قرار می گیرند. از طرف دیگر میتواند بد باشد چون سیاستهایی که بوسیله افرادی که از نظر تکنولوژی آگاه نیستند ایجاد شده اند غالباً با واقعیت فاصله دارند: نگاهی به «مناظره پشت درها برای کدگذاری پروتکلها» ببندازید

بعضیها میگویند سازمانهایی که از داده استفاده میکنند باید مدل کمیته بازنگری سازمانی (IRB) از صنعت زیست پزشکی را بپذیرند. متأسفانه با اینکه مدل کمیته بازنگری سازمانی جنبه‌های مثبتی دارد، اما نمیتواند رویکرد قابل اطمینانی باشد. کمیته های بازنگری سازمانی پیچیده هستند و نمیتوانند چابک باشند. خیلی سخت است که یک کمیته بازنگری سازمانی با ایده‌ها و تکنولوژیهای جدید وفق داده شود. به همین دلیل است که در دولت اوپاما بروزرسانی مدل قانون عمومی¹⁰² برای کسب رضایت کاربر به طوری که بتواند با تکنولوژیهای دیجیتالی سازگار باشد و قابلیت داده کاوی نیز داشته باشد، حدود ۸ سال طول کشید.

⁹² The Nuremberg Code

⁹³ National Research Act

⁹⁴ 1975 Declaration of Helsinki

⁹⁵ Institutional Review Board

⁹⁶ Common rule : The **Common Rule** is a 1981 rule of ethics in the [United States](#) regarding biomedical and behavioral research involving human subjects.

⁹⁷ The US Federal Trade Commission

⁹⁸ The Nuclear Regulatory Commission

⁹⁹ The Federal Food and Drug Administration

¹⁰⁰ The Consumer Finance Protection Bureau

¹⁰¹ General Data Protection Regulation

¹⁰² Common Rule

ساختن آینده مان

مدتی است که از مسائل اخلاقی که به دنبال استفاده یا سو استفاده از داده ایجاد می‌شوند آگاهییم. اعتراضات عمومی علیه فیسبوک نهایتاً فروکش خواهد کرد، اما مشکلاتی که باعث این اعتراضات شدند خیر. ما با آینده‌ای روبرو هستیم که در آن اکثر اتومبیلها خودکار هستند و ما با روباتهای صحبت خواهیم کرد که نمیشود آن‌ها را از انسان‌ها متمایز کرد و به جایی خواهیم رسید که دستگاهها به همه مکالمات ما گوش میدهند و آماده اند تا به ما پیشنهادهای سازنده در مورد هر چیزی از رستوران و دستور پخت غذا گرفته تا دستورات پزشکی بدهند. نتیجه این پیشرفتهای میتواند شگفت انگیز باشند و یا در حد کابوس وحشت‌انگیز باشد.

آن روز، روز حساب پس دادن داده است. تصویر آینده ما بسیار وابسته به آنچه که ما در چند سال آینده انجام خواهیم داد می باشد. لازم است که اخلاق را در همه جنبه‌های آموزش فنی و فرهنگ سازمانی شرکت دهیم. باید به افراد آزادی توقف تولید را در شرایط ضروری و ابراز نگرانی هایشان در صورتی که رفع نشوند را بدهیم. باید تنوع و اخلاق را در تصمیمات استخدامی لحاظ نماییم و ممکن است لازم باشد برای حمایت از منافع افراد و کل جامعه قوانینی را در نظر بگیریم.

در کنار همه اینها، در خصوص اخلاق صحبت کنید. در مقاله «زمان آن است که سر میز شام در مورد اخلاق داده صحبت شود»¹⁰³ ناتالی ایوانز هریس¹⁰⁴ و سایر نویسندگان این مقاله نوشته‌اند که «باید یک مکالمه عمیق در مورد مسئولیتهای فردی و جمعی مان در مورد استفاده اخلاقی از داده داشته باشیم». این بهترین چیزی است که میتوانید برای بیشتر شدن اخلاق داده انجام دهید. در جلسات نشستها، ناهار و حتی شام در مورد آن صحبت کنید. اگر با اخلاق زندگی نکنید و نفس نکشید، امضا یک سوگندنامه داده یا موافقت با یک دستورالعمل کمک چندانی نمیکند. اما وقتی که با آن زندگی کنید شروع به فکر کردن در مورد کدهایی که مینویسید، مدلی که میسازید و برنامه‌ای که تولید میکنید خواهید نمود. تنها راه ساختن یک اخلاق فرهنگی زندگی کردن در آن است. تغییرات به صورت جادویی اتفاق نخواهد افتاد و آسان هم نیست، اما لازم است.

ما میتوانیم آینده‌ای که میخواهیم در آن زندگی کنیم را بسازیم و یا میتوانیم یک کابوس را بسازیم. انتخاب با ما است.

فصل پنجم: مورد پژوهی

برای اینکه بتوانیم به طور جدی در باره اخلاق داده فکر کنیم، نیاز به موارد پژوهشی ای داریم تا بتوانیم در موردشان بحث و تبادل نظر کنیم و در حالیکه با جهان واقعی درگیر هستیم به تفاهم برسیم. موارد پژوهشی خوب به ما این فرصت را میدهند که قبل از اینکه در زندگی واقعی به مشکلی بر بخوریم برایش راه حلی اندیشیده باشیم و پژوهشها به ما نشان میدهند که مسائل اخلاقی ساده نیستند. آن‌ها جوانب متعددی دارند و غالباً یک راه حل خاص ندارند و به ما کمک میکنند که بدانیم که موقعیتهایی که پرسشهای اخلاقی را به دنبال نداشته باشند، نادر خواهند بود.

مرکز سیاستگذاری فناوری اطلاعات پرینستون¹⁰⁵ و مرکز ارزش های انسانی¹⁰⁶ چهار مورد بی نام برای ترفیع بحثهای اخلاقی ایجاد نموده اند(موارد بیشتر در پروسه تدوین هستند و ممکن است در زمانیکه این را میخوانید در دسترس باشند). اولین مورد از این پژوهشها برنامه سلامت خودکار¹⁰⁷ است که در مورد اپلیکیشنی روی ساعت‌های هوشمند صحبت میکند که به بزرگسالانی که به تازگی درگیر دیابت شده‌اند کمک میکند. این مورد مباحثی مانند پدرا نه رفتار کردن، رضایت و حتی انتخاب زبان را به دنبال دارد. آیا درست است که بیماران را در جهت رفتارهای سالم تر سوق بدهیم؟ در مورد اداره خودکار

¹⁰³ It's time for data ethics conversations at the dinner table

¹⁰⁴ Natalie Evans Harris

¹⁰⁵ Princeton's Center for Information Technology

¹⁰⁶ Center for Human Values

¹⁰⁷ Automated Healthcare App

گروههای کاربری برای تأکید بر اطلاعات دقیق علمی چه کنیم؟ و با اقلیتی که به درمانها پاسخ مناسب ندهند چه کار کنیم؟ آیا میتواند مشکل از زبانی باشد که برای توضیح درمان از آن استفاده شده است؟

مورد پژوهشی دیگر، تشخیص صدای پویا است¹⁰⁸ که اپلیکیشنی برای تشخیص صدا در نظر گرفته شده است و مسائلی مانند حریم، زبان و حتی جنسیت را به دنبال دارد. توسعه دهنده ها تا چه حد باید به مبحث شناسایی صدماات احتمالی که ممکن است توسط برنامه ایجاد شود، وارد شوند؟ نرخ خطای قابل پذیرش برای برنامه‌ای که احتمال صدمه زدن دارد چقدر است؟ چگونه یک برنامه تشخیص صدا میتواند گونه های مختلف لهجه و گویش را اداره نماید و مسئولیت توسعه دهندگان وقتی یک ابزار آزمایشی کوچک بوسیله یک شرکت بزرگ که قصد تجاری سازی آن را دارد خریداری میشود، چیست؟

مورد پژوهشی دیگر، مدرسه بهینه سازی¹⁰⁹ است که هدفش یافتن کودکان در معرض خطر در مدرسه است. بازهم حریم و زبان مورد بحث میباشند. همچنین این بحث مطرح می‌شود که تصمیم استفاده از داده چگونه گرفته شود. چه کسی آن تصمیمات را بگیرد و چه کسی از آن‌ها مطلع شود؟ چه پیامدهایی بوجود خواهد آمد وقتی مردم بفهمند که از داده آن‌ها چگونه استفاده شده است؟ و چگونه نتایج یک آزمایش را تفسیر خواهید نمود؟ تحت چه شرایطی میتوانید بگویید که یک آزمایش داده نتایج آموزشی را بهتر نموده است؟

مورد آخر، چت باتهای ناظر بر قانون¹¹⁰ می باشند. مسائل مورد بحث در این مورد حد و مرز بین امنیت و آزادی، اجبار، اراده و پاسخگویی و تسلیم در برابر قوانین بین‌المللی میباشد.

هیچیک از این مباحث ساده نیستند و پاسخهای درست بسیار کم هستند (اگر وجود داشته باشند). برای مثال ساده است که در مقابل مشاهده رفتار پدران یک برنامه پزشکی واکنش نشان دهیم اما هدف اینچنین برنامه‌ای تشویق بیمار به تبعیت از برنامه‌های درمانی است. ساده است که در مقابل مانیتور کردن دانش آموزان در مدرسه عمومی اعتراض نماییم. اما دانش آموزان خردسال هستند و طبیعتاً مدرسه‌ها اطلاعات شخصی محرمانه زیادی را میداند. مرز بین آنچه قابل پذیرش است و نیست کجاست؟ چیزی که مهم است رسیدن به جواب درست برای هر موضوع نیست، بلکه هدف، اطمینان از درک و فهم موضوع می‌باشد و اینکه ما بدانیم چه چیزهایی را فدای چه چیزهایی میکنیم. چیزی که مهم است این است که ما داریم در خصوص توضیح مسایل اخلاقی عمل میکنیم و این عمل را در شغلان هم بیاوریم. این چیزی است که مورد پژوهی در اختیار ما قرار میدهد.

¹⁰⁸ Dynamic Sound Identification

¹⁰⁹ The Optimizing School

¹¹⁰ Law Enforcement Chatbots

درباره نویسندگان

مایک لوکیدز¹¹¹ قائم مقام استراتژی محتوا در اورایی مدیا¹¹² است. او کتابهای مورد توجه زیادی را در موضوعات فنی ویرایش کرده است که البته شامل برنامه نویسی ویندوز نمی شود. او به طور خاص به زبانهای برنامه نویسی، یونیکس¹¹³ (و آنچه این روزها بر یونیکس می گذرد) و مدیریت شبکه و سیستم علاقه مند است. مایک نویسنده کتاب بهینه سازی کارایی سیستم¹¹⁴ و یکی از نویسندگان کتاب ابزارهای یونیکس¹¹⁵ است. اخیرا هم درگیر داده، تحلیل داده ها و زبانهای مثل R، Mathematica و Octava شده است و روی اینکه چطور می توان کتابها را تبدیل به یک شبکه اجتماعی کرد فکر می کند. از طریق شناسه mikeloukids@ در توییتر و لینکدین می توانید با مایک در ارتباط باشید.

هیلاری میسون¹¹⁶ مدیر یادگیری ماشین در کلودرا¹¹⁷ است و قبل از آن هم فست فرووارد لبز¹¹⁸ را تاسیس کرده است که یک شرکت مشاوره و پژوهش در زمینه یادگیری ماشین کاربردی است. این شرکت در سال 2017 توسط کلودرا خریداری شد. او همچنین به عنوان دانشمند داده در اکسل پارتنرز¹¹⁹ فعالیت می کند و عضو هیئت مدیره موسسه آنتا بورگ¹²⁰ است. و از هم بنیانگذاران HackNY.org است که یک موسسه غیرانتفاعی است که به دانشجویان مهندسی در پیدا کردن فرصتها و موقعیت های کاری در نیویورک کمک می کند. عضو شورای مشاوره فنی بلومبرگ¹²¹ بوده است و در بیتلی¹²² از دانشمندان ارشد داده. از طریق شناسه hmason@ می توانید در توییتر با او در تماس باشید.

دی.جی. پاتیل¹²³ اولین دانشمند داده ارشد دولت آمریکا که استراتژی ملی برای داده ها را برای امریکا رهبری کرد. در حال حاضر در حال ساختن Devoted Health است، یک شرکت بهداشتی که ضیفه خود می داند از تمام اعضای خود مثل یک خانواده مراقبت کند. وی همچنین عضو ارشد مدرسه کندی هاروارد و مشاور ونراک پارتنرز است. قبلا به عنوان قائم مقام در relatedIQ که توسط [salesforce.com](https://www.salesforce.com) خریداری شد فعالیت می کرد. وی همچنین مدیر محصولات داده ای و امنیت اطلاعات در لینکدین بود. از طریق شناسه dpatil@ در توییتر می توانید با او در تماس باشید.

Data professionals

Escalation process

Data ethics: اخلاق داده

¹¹¹ Mike Loukides

¹¹² O'reilly Media

¹¹³ Unix

¹¹⁴ System Performance Tuning

¹¹⁵ Unix Power Tools

¹¹⁶ Hillary Mason

¹¹⁷ Cloudera

¹¹⁸ Fast Forward Labs

¹¹⁹ Accel Partners

¹²⁰ Anita Borg Institute

¹²¹ Mayer Bloomberg

¹²² Bitly

¹²³ DJ Patil

رضایت آگاهانه : Informed consent

پاپ آپ : Pop up

بی‌طرفی : Bias

شرکتهای تجارت آنلاین : Consumer Internet Service

باید کلمات خاص مثل بایاس را همه جا به یک شکل ترجمه کنیم که خواننده متوجه بشه داریم راجع به چی حرف میزنیم

را قانون بنویسیم یا اصول ؟ Principal

کارهای باقی مانده :

ریویوی کلی

لینک دادن لینک ها

نوشتن یک معرفی نامه

خواندن بخش بخش و درست کردن انشا