BRANDAFARIN JOURNAL OF MANAGEMEMNT
Volume No.: 2, Issue No.: 21, Dec 2021
Print ISSN: 2717-0683

Evaluate the use of artificial intelligence in decision-making systems of government agencies

Homayoun Monazami

DBA graduate of Tarjoman Oloom Higher Education Institute

Abstract

of artificial intelligence (AI) based software, due to the technical complexity of the software artifact and, often, its embedding in complex sociotechnical processes. Recent advancesin machine learning (ML) enabled by deep neural networks has exacerbated the challenge of evaluating such software due to the opaque nature of these ML-based artifacts.
A key related issue is the (in)ability of such systems to generate useful explanations of their outputs, and we argue that the explanation and evaluation problems are closely linked.
The paper models the elements of a ML-based AI system in the context of public sector decision (PSD) applications involving both artificial and human intelligence, and maps these elements against issues in both evaluation and explanation, showing how the two are related. We consider a number of common PSD application patterns in the light of our model, and identify a set of key issues connected to explanation and evaluation in each case. Finally, we propose multiple strategies to promote wider adoption of AI/ML technologies in PSD, where each is distinguished by a focus on different elements of our model, allowing PSD policy makers to adopt an approach that best fits their context and concerns.

Keywords: Artificial intelligence, public administration, public sector decision making

ارزیابی استفاده از هوش مصنوعی در سیستم های تصمیم گیرنده ارگان های دولتی

همایون منظمی

دانش آموخته DBA موسسه آموزش عالی ترجمان علوم

چکیده

"ارزیابی" همیشه یک چالش اساسی در توسعه نرم افزارهای مبتنی بر هوش مصنوعی (AI) به دلیل پیچیدگی فنی نرم افزار و فرآیندهای پیچیده اجتماعی – فنی موجود درآن، بوده است. پیشرفت های اخیر در یادگیری ماشین (ML) که توسط شبکه های عصبی عمیق امکان پذیر شده است ، چالش ارزیابی چنین نرم افزارهاییی را به دلیل ماهیت مبهم این محصولات مصنوعی مبتنی بر ML ، تشدید کرده است. یک موضوع مهم کلیدی ، توانایی (ناتوانی) این سیستم ها در تولید توضیحات مفید در مورد خروجی هایشان است و ما معتقدیم که مسائل توضیح و ارزیابی به طور تنگاتنگی با هم مرتبط هستند. این مقاله عناصر یک سیستم AI مبتنی بر ML را در زمینه برنامه های تصمیم گیری بخش دولتی (PSD) شامل هوش مصنوعی و انسانی مدل سازی می کند ، و در ارزیابی و توضیح، این عناصر را بر خلاف مسائل ترسیم می کند و نحوه ارتباط این دو را نشان می دهد. ما با توجه به مدل خود تعدادی مدل کاربرد PSD متداول را در نظر می گیریم و مجموعه ای از موضوعات اصلی مرتبط با توضیح و ارزیابی را در هر مورد شناسایی می کنیم. در نهایت ، ما چندین استراتژی را برای ترویج پذیرش گسترده تر فناوری های AI / ML در PSD پیشنهاد می کنیم ، به طوری که هر کدام بر اساس تمرکز بر عناصر مختلف مدل ما تفکیک می شوند ، و به سیاست گذاران PSD اجازه می دهد رویکردی که متناسب با زمینه و نگرانی های آنها باشد، انتخاب کنند.

کلیدواژگان: هوش مصنوعی، مدیریت دولتی، تصمیم گیری بخش دولتی

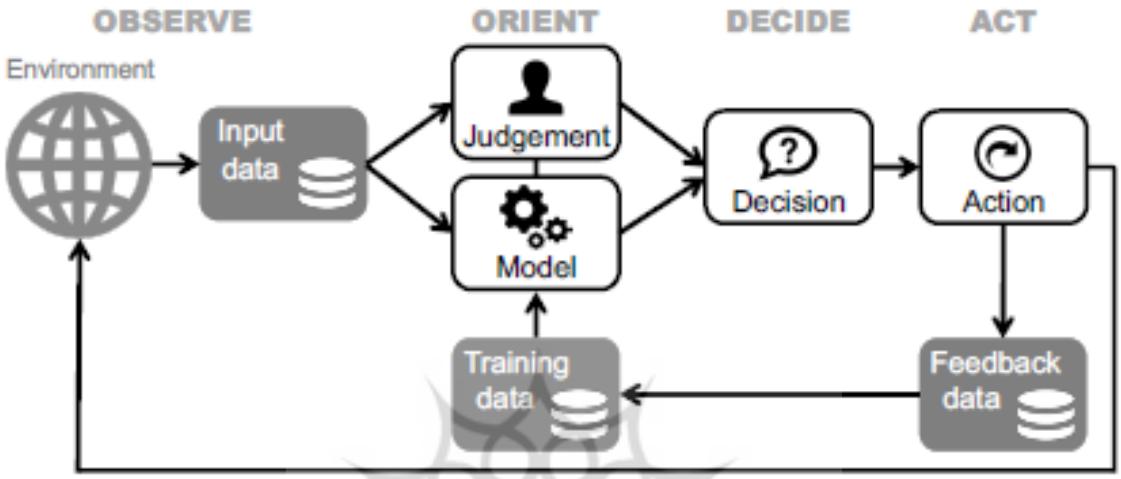
مقدمه

ارزیابی[1] در توسعه نرم افزار مبتنی بر هوش مصنوعی (AI[2]) یک مساله دشوار است و همیشه هم، همین طور بوده است. برای هر سیستم نرم افزاری، ارزیابی از درستی یابی[3] و اعتبار سنجی[4] (IEEE, 1990) تشکیل می شود، به طور محاوره ای به صورت ذیل تعریف می شود: درستی یابی بر "درست سازی سیستم" متمرکز است (یعنی اطمینان حاصل نمودن از رعایت مشخصات فنی آن)؛ اعتبار سنجی بر "ساخت سیستم درست" تمرکز دارد (یعنی، اطمینان حاصل نمودن از اینکه این سیستم، نیازهای ذی نفعان خود را تامین میکند). هر دو برای یک سیستم مبتنی بر هوش مصنوعی، دشوار هستند: درستی یابی، چون پیچیدگی فنی مصنوع نرم افزاری بر تکنیک های درستی یابی نرم افزاری سنتی پافشاری دارد (اُکیف و اُلری، 1993)؛ اعتبار سنجی، چون سیستم های مبتنی بر هوش مصنوعی، معمولا به صورت بخشی از یک سیستم تصمیم گیری پیچیده عمل می کنند که شامل

---

[1] Evaluation
[2] Artificial inteligence
[3] verification
[4] validation

هم عناصر نرم افزاری و هم انسانی می شود، سنجش و اطمینان از اثرات مد نظر، بسیار دشوار است (کامینگز، 2014). سال های اخیر، شاهد پیشرفت های سریعی در قابلیت نرم افزار هوش مصنوعی بوده است که عمدتا به خاطر پیشرفت ها در یادگیری ماشین (ML[5]) با استفاده از شبکه عصبی چند لایه (موسوم به عمیق) بوده است (لیکان، بنژیو و هینتون، 2015). در عین حال، این امر به علاوه، پیچیدگی فنی مصنوعات نرم افزاری را افزایش داده است و حتی به چالش های بیشتر در درستی یابی و اعتبار سنجی منجر شده است (گودفلو، مک دنیل و پیپرنات، 2018).



شکل 1: حلقه تصمیم گیری انسان + ماشین ، براساس OODA

از ابتدا در توسعه گسترده نرم افزار مبتنی بر هوش مصنوعی، به رسمیت شناخته شد – در طول عصر "مهندسی دانش" دهه 1980 (باچنان و شرتلیف، 1984) – که تسهیلات توضیح، رابطه نزدیکی با درستی یابی و اعتبار سنجی داشتند، زیرا آنها، قابل بررسی بودن ضروری را ارائه می دهند: 1) به توسعه دهندگان سیستم نرم افزاری، برای کمک به اشکال زدایی[6]، یک عنصر مهم در درستی یابی ("درست سازی سیستم") کمک می کند؛ 2) به کاربران نهایی برای افزایش اعتماد به سیستم، یک عنصر مهم در اعتبار سنجی، کمک می کند ("ساخت سیستم صحیح"). کار اولیه در توضیح برای سیستم های مبتنی بر هوش مصنوعی، دو نوع درخواست توضیح متناظر را شناسایی نمود:

چطور؟ یک "منطق" را برای استدلال سیستم، درخواست نماییم (یعنی، چرا X را باور دارید؟").

یک دسته قابل توجه از تحقیقات و پیشرفت ها در زمینه توضیح نرم افزاری هوش مصنوعی، درستی یابی و اعتبار سنجی از این تمایز مهم، نشات می گیرد، پس از اینکه درک شد، ذی نفعان مختلف به انواع بسیار متفاوتی از توضیحات نیاز دارند، بسته به اینکه آیا علاقه شان، در اصل به درستی یابی است یا اعتبار سنجی (جکسون، 1999). یاداوری کننده های این موضوع، هنوز امروز ساخته می شوند، مثلا (کرچ، 2017، تامست و دیگران، 2018). در حقیقت، توضیح، یکبار دیگر به عنوان یک مساله بحرانی در عصر یادگیری ماشین گرای کنونی هوش مصنوعی پدید آمده است، گرچه، عدم وضوح در زمینه واژگان وجود دارد

[5] Machine learning
[6] debug

(لیپتون، 2016). عموما، قدری اتفاق نظر وجود دارد که دو رویکرد وجود دارد: 1) شفافیت به منظور توضیح خروجی های سیستم بر حسب طرز کارهای داخلی سیستم است (یعنی شفافیت در یک مفهوم فنی، همانند تجسم های فکری فعالسازی های شبکه عصبی)؛ 2) توضیحات پسین[7]، به منظور توجیه خروجی ها، بر حسب توجیهات عقلی طرز کارهای سیستم به جای تلاش برای نشان دادن طرز کارهای واقعی است (مثلا "توضیح توسط نمونه) بر حسب انتخاب نمونه های قبلا مشاهده شده مشابه از طریق مثال).

در این مقاله، ما بر هوش هوش مصنوعی/ یادگیری ماشینبرای کاربردهای تصمیم گیری بخش عمومی (PSD[8]) تمرکز داریم که در آنجا ارزیابی و توضیح، به دلایل متعدد، از اهمیت بسزایی ، برخوردارند. به طور ویژه ای ارزیابی اثر کاربرد هوش مصنوعی برای تصمیم گیری بخش عمومی، چالش برانگیز است زیرا سیستم معمولا در یک فرایند اجتماعی- فنی تعبیه خواهد شد که ممکن است، سنجش و اعتبار سنجی اثراتش، دشوار باشد. به علاوه، توضیح در چنین کاربردهایی، رابطه نزدیکی با مسائل ذی حسابی دارد (دایکوپولوس، 2016) که به طور خاصی در مورد یکان هایی که بودجه شان به طور عمومی تامین می شود و فرایندهای تصمیم حاد هستند. نهایتا، در اکثریت قریب به اتفاق موارد، تصمیمات و اعمال توسط حامیان بشر در کاربردهای تصمیم گیری بخش عمومی اجرا می شوند که استانداردهای بالایی را برای ارزیابی و توضیح در زمینه همکاری ماشین- انسان قرار می دهند (تروین، 1995).

این مقاله به صورت ذیل، برنامه ریزی می شود: ما برای ایجاد چارچوبی برای بحث بعدی، با شناسایی عناصر حلقه تصمیم مبتنی بر هوش مصنوعی انسان + ماشین کار خود را آغاز کرده و بررسی می کنیم، ارزیابی (درستی یابی و اعتبار سنجی) و توضیح چطور به معلوم ها/مجهول های موجود در حلقه می پردازند؛ آنگاه انواع رایج کاربردهای هوش مصنوعی برای تصمیم گیری بخش عمومی را بررسی می کنیم و آنها را در چارچوب خودمان، ترسیم می نماییم؛ آنگاه بستر وسیعتر توسعه هوش مصنوعی را برای تصمیم گیری بخش عمومی بر حسب کیفیت داده ها، نیرومندی، همکاری ماشین- انسان، مالکیت دارایی و اینکه "چه چیزی در هر مورد اشاره کننده به کانون های مهم بر حسب چارچوب اولیه ما درست عمل می کند" را بررسی می کنیم؛ نهایتا، راهبردهای متعددی برای تصمیم گیری بخش عمومی هوش مصنوعی گرا را پیشنهاد می دهیم، بار دیگر، چارچوب خود را استنباط می نماییم:"ماموریت گرا"، "داده گرا"، "کار گرا" و "شواهد گرا".

عناصر یک حلقه انسان + تصمیم هوش مصنوعی

برای بیان بحث بعدی، شکل 1، مدل مفهومی یک چرخه تصمیم شامل یک سیستم هوش مصنوعی مبتنی بر ML را نشان می دهد که با یک تصمیم گیرنده انسانی کار می کند. این مدل از حلقه (مشاهده ، جهت یابی، تصمیم گیری، عملکرد) OODA بوید اقتباس شده است (بوید، 1995). عناصر اصلی این حلقه، به صورت ذیل هستند:

داده های ورودی: جمع آوری شده از محیط و متشکل از مشاهداتی که به صورت ورودی برای فرایند تصمیم گیری، لازم است.

مدل: مدل ایجاد شده توسط یک الگوریتم ML از یک مجموعه از داده های آموزشی؛ برای استفاده در تصمیم گیری، این مدل به صورت بخشی از یک سیستم نرم افزاری گسترش می یابد.

---

[7] Post hoc: تعقیبی،پس از این

[8] public sector decision

داده های آموزشی: بکار رفته برای ساخت مدل از طریق یک الگوریتم ML، متشکل از (برای یادگیری نظارت شده) یک مجموعه معمولا بزرگ از جفت های ورودی- خروجی توصیف کننده موارد تصمیم گیری قبلی.

قضاوت: برای سادگی بیشتر به عنصر انسانی فرایند تصمیم گیری به عنوان "قضاوت" اشاره می کنیم که ترکیبی از دانش (بویژه رموز کار ضمنی)، تجربه و خرد را بازتاب می نماید (کالینز، 2010).

تصمیم گیری: صورت گرفته در همکاری توسط انسان + ماشین در سطحی از اتوماسیون از یک کران، ماشین، برای آگاهی بخشی به یک تصمیم گیرنده انسانی در کران دیگر، پیشنهادات خود را ارائه می دهد، انسان دارای درجه ای از "حق وتو[9]" یا سرپرستی در زمینه تصمیم گیرنده ماشینی است (کامینز، 2014).

عملکرد: صورت گرفته توسط انسان ها، ماشین ها یا ترکیبی از هر دو که به فیدبک و تغییرات محیط منجر می شود.

داده های فیدبک: نتایج اقدام نمودن را توصیف می نماید (شاید مثبت یا منفی باشد) که ممکن است به مجموعه داده آموزشی بازخورد پیدا کند (آیا این بازخورد یا فیدبک بیانگر نیازی برای آموزش مجدد است).

عناصر داده های مدل به صورت کادرهای خاکستری نشان داده شده اند؛ عناصر عملیاتی، به صورت کادرهای سفید نشان داده شده اند.

بر حسب OODA، جمع آوری داده ها از محیط، از مرحله مشاهده تشکیل می شود و ممکن است، شکل ادراک فیزیکی را به خود بگیرد (مثلا، گرفتن عکس از طریق یک دوربین) یا انتقال اطلاعات از طریق اسناد (مثلا داده های جمع اوری شده از طریق فرم ها). مدل مبتنی بر ML، به طور کلی روی این داده های ورودی به عنوان "جهان بینی" اش عمل خواهند کرد؛ ممکن است تصمیم گیرنده بتواند، محیط، هم چنین دستیابی به داده های ورودی را مستقیما درک کند (ممکن است به شکلی متفاوت از شکلی باشد که ماشین به آن دستیابی دارد). مدل های متعدد مرحله بوید گرا به مدل مبتنی بر ML و قضاوت انسانی تقسیم می شود. مراحل تصمیم و عمل، ضرورتا، در اینجا همانند OODA یکسان هستند، با تفاوتی که مدل بوید برای فیدبک متوالی از محیط در طول مرحله تصمیم میسر است- که برای سادگی ما آنها را حذف می نماییم- و مدل ما، داده های فیدبک صریح را به صورت خروجی از مرحله عملکرد نشان می دهد. مدل OODA هم چنین "راهنمایی و کنترل" بین مرحله جهت یابی و هم مراحل مشاهده و هم عملکرد را نشان می دهد که بار دیگر، برای سادگی حفظ می شوند.

اثبات شده است، توضیح یک عنصر کلیدی در مرحله جهتیابی OODA است زمانیکه به تصمیم گیری هدف گرا می پردازیم (آها 2018). در حقیقت، ما توضیح را به عنوان عنصر مهمی در تعامل بین انسان و عوامل ماشینی در این مدل تلقی می کنیم (نشان داده شده از طریق اتصال خطی این دو عنصر در شکل). بر حسب نیاز برای توضیحات، چندین ظرافت بالقوه وجود دارد که صریحا در شکل 1 نشان داده شده اند، برخی از آنها، در اثر تامست و دیگران (2018) مورد بررسی قرار می گیرند. مثلا، تصمیم گیرنده انسانی ممکن است، مستقیماً با سیستم های مبتنی بر هوش مصنوعی تعامل نداشته باشد، بلکه در عوض ممکن است از طریق یک اپراتور، ارتباط برقرار نماید. در این صورت، اپراتور و تصمیم گیرنده ممکن است به اشکال متفاوتی از توضیح، نیاز داشته باشند. به طور مشابه، انسان هایی که عملی را انجام می دهند، ممکن است به اشکال متفاوتی از توضیح، بار دیگر، نیاز داشته باشند. نهایتا، ممکن است، انسان هایی در محیط وجود داشته باشند که تحت تاثیر عملکردی قرار داشته باشند که

[9] veto

ممکن است دارای یک "حق توضیح" باشند همانطور که اخیرا در قانون اروپا محفوظ شده است (گودمن و فلکسمن 2016). بعدا به این بحث باز می گردیم.

اکنون با توجه به عناصر شکل 1 به ارزیابی باز می گردیم، اثبات می کنیم، هدف ارزیابی، در نظر گرفتن فضای معلومها/مجهول ها به صورت ذیل است:

معلوم های معلوم: مواردی هستند که خالقان سیستم هوش مصنوعی، آنها را می شناسند، مدل باید در حدود داده های آموزشی، آنها را بشناسد. آنها از طریق درستی یابی، قابل آزمایش هستند (و در جای لازم، مجددا آموزش داده می شوند) و برای اهداف اشکال زدایی از طریق روش های نوع شفافیت، قابل شرح هستند.

مجهول های معلوم: جستجوهایی هستند که خالقان سیستم هوش مصنوعی انتظار دارند، سیستم بتواند آنها را انجام دهد، یعنی چیزهایی که از آموزش شان، "قابل پیش بینی" هستند. آنها از طریق اعتبار سنجی، قابل آزمایش هستند و برای اهداف اعتماد کاربر از طریق روش های توضیح پسین، قابل توضیح هستند. پیش بینی های ترکیب شده با قضاوت انسانی، در تصمیمات و اعمال درونگذاری می شوند و در نهایت از طریق داده های فیدبک اعتبار سنجی می شوند (ممکن است مستلزم آموزش مجدد باشند).

معلوم های مجهول: از چشم انداز مدل، چیزهای خارج از قلمروی ماشین هستند اما در قلمروی دانش تصمیم گیرنده و قضاوت انسانی قرار دارند. توضیحات (شفاف یا پسین) ممکن است برای آشکار نمودن این مجهولات ماشینی، لازم باشند. مجموع توانایی انسان + ماشین برای پرداختن به آنها باید از طریق داده های فیدبک اعتبار سنجی شوند، به طور بالقوه با آموزش مجدد سیستم در مورد فیدبک منفی دنبال می شود.

مجهول های مجهول: چیزهایی هستند که خارج از قلمروی هم مدل و هم دانش و قضاوت انسان، قرار دارند. به طور محاوره ای، اینها، " تله ها[10]" هستند. نیرومندی سیستم های مبتنی بر هوش مصنوعی برای مجهول های مجهول، به صورت یک مساله کنونی مهم، پرچم دار شده است، گرچه روشهای متعددی برای کاهش آنها، شناسایی شده اند (دیتریچ، 2017)؛ اعتبار سنجی مجموع انسان + ماشین به ارزیابی روش های کاهش منتخب نیاز دارند.

انواع مساله تصمیم گیری بخش عمومی

مولگان (مولگان، 2017) شش مرحله را در فرایند تصمیم گیری بخش عمومی شناسایی می نماید:

1) بیان سوالات مورد توجه؛
2) شناسایی مسائلی که ممکن است، متمایل به عملکرد باشند؛
3) ایجاد گزینه ها برای درستی یابی؛
4) گزینه های موشکافی/سنجش؛
5) تصمیم گیری (انتخاب یک گزینه)؛
6) قضاوت در مورد اینکه آیا درست عمل می کند.

[10] Gotcha: دام

این مساله ممکن است هم به عنوان یک فرایند تصمیم و هم یک فرا- فرایند برای انتخاب موارد مسائل تصمیم گیری بخش عمومی تلقی شود. به این ترتیب، حلقه تصمیم را در بخش قبلی به دو روش، تقسیم می کند. برای در نظر گرفتن آن اولا به صورت یک فرا فرایند: مرحله 1 تا 3 را در نظر می گیرند، کدام ابعاد محیط "در قلمروی" سوالات هستند، آیا داده های مناسب (ورودی و آموزش)، قابل جمع اوری هستند، آیا الگوریتم های ML مناسب برای تولید مدل های قوی، در دسترس هستند و تعادل بین ماشین و هوش مصنوعی در پرداختن به این سوالات مد نظر قرار گرفته می شوند. مرحله 4، هم موافقین و هم مخالفین روش های ML و به طور مهمی، سطوح اتوماسیون جایگزین بین ماشین و انسان را در نظر می گیرند. مرحله 5، شامل اجرا و درستی یابی یک رویکرد منتخب می شود. مرحله 6، به مسائل وسیعتر درستی یابی و فیدبک می پردازد.

با در نظر گرفتن شش مرحله به عنوان یک فرایند تصمیم، مراحل 1 و 2، مشاهده را در بر می گیرد، مراحل 2-4، جهتیابی را در بر می گیرد، مرحله 5، تصمیم و عملکرد را در بر می گیرد و مرحله 6، فیدبک را مد نظر قرار می دهد. در این دیدگاه، ML، زمانی قابل کاربرد می شود که می تواند برای کمک به بیان سوالات، شناسایی مسائل و تولید گزینه ها بکار رود. این کاربرد، یک نوع کاربرد داده کاوی کلاسیک/ کشف دانش است، با این نوع کاربرد به طور کلی، توضیح مدل آموخته شده در شرایطی لازم است که برای یک کاربر نهایی، معنادار و مفید (یعنی انتقال دانش کشف شده)و حیاتی باشد (براتکو، 1997).

شش "الگوی" کاربرد تحلیل داده تصمیم گیری بخش عمومی که همگی روی مدل بخش قبلی ترسیم می شوند- توسط اداره عملکرد و ذی حسابی نیو ارلینز توصیف شده است:

الف. "یافتن سوزن در یک انبار کاه": موارد غیر عادی را شناسایی کنید، مثلا با آموزش یک مدل پیش بینی کننده در زمینه موارد غیر عادی گذشته؛

ب) **"اولویت بندی کار برای اثر"**: دسته بندی موارد بر حسب بالاترین خطر یا بالاترین ارزش؛

ث) **"ابزارهای هشدار اولیه"**: مسائل را در مراحل آغازین پیش از افزایش تدریجی کشف نمایید، مثلا از یک الگوی شکایات مکرر؛

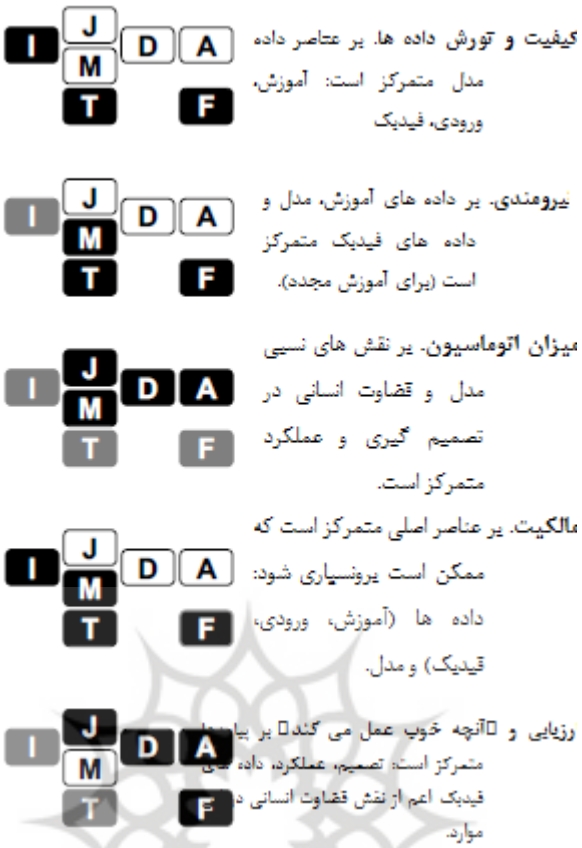
د) **"تصمیمات بهتر سریعتر"**: کیفیت تصمیم و بجا بودن بهتر هدایت شده از طریق استفاده بیشینه داده های موجود (موارد گذشته برای آموزش و داده های ورودی در زمینه موقعیت کنونی)؛

ذ)**"بهینه سازی تخصیص منبع"**: بازده سازمانی بهبود یافته با پتانسیل کاهش هزینه با استفاده از means/ends[11] analysis است؛

ر) **"آزمایشگری برای آنچه خوب عمل می کند"**: آزمایش الف و ب در زمان اجرای سازمانی با فیدبک دینامیکی.

برای سه مورد اول، داده های آموزشی مناسب، عنصر کلیدی محسوب می شود. برای چهارمی، کیفیت تصمیم گیری را از طریق همکاری انسان + ماشین، به حداکثر می رساند در حالیکه پس افتادگی های بین داده های ورودی و عملکرد و بین عملکرد و استفاده از داده های فیدبک را به حداقل می رساند. دو مورد آخر بر حلقه به صورت یک کل، تمرکز دارند.

[11] نوعی روش استدلال که در یک کوشش برای کاهش تفاوت ها از نقطه شروع تا هدف به عقب و جلو نظر می اندازد.



شکل 2. عناصر اصلی مدل (نشان داده شده به رنگ سیاه) در تاکید برای هر مساله تصمیم گیری بخش عمومی .

بحث قبلی، حاکی از این مساله است که "مالک" فرایند تصمیم گیری بخش عمومی، یک یگان دولتی یا بخش عمومی است. یک دیدگاه دیگر تصمیم گیری بخش عمومی، به منظور قدرت بخشیدن به شهروندان برای بهبود توانایی خود برای تصمیم گیری و عملکرد در رابطه با خدمات عمومی و جامعه مدنی است (مولگان، 2017).

هوش مصنوعی برای تصمیم گیری بخش عمومی : مسائل

این بخش، چندین مساله مهم مستلزم توجه در توسعه رویکردهای هوش مصنوعی/یادگیری ماشین برای تصمیم گیری بخش عمومی را مورد بحث قرار می دهد. شکل 2، عناصر این حلقه را مورد تاکید قرار می دهد که مورد تمرکز اصلی برای هر یک از این مسائل قرار دارند. در هر صورت، این مدل به صورت یک نسخه "جدول دوره ای" انتزاعی شکل 1 با عناصر داده به رنگ خاکستری، عناصر عملیاتی به رنگ سفید و عناصر تمرکز (داده ها یا عملیاتی) در هر صورت به رنگ مشکی نشان داده شده است.

کیفیت و تورش داده ها. اینجا، تمرکز بر عناصر داده ای حلقه است: ورودی، آموزش، فیدبک. مسائل مهم عبارتند از: 1) کیفیت مجموعه آموزش با توجه به قلمروی مجهول های معلوم (درستی یابی) و مجهول های معلوم (اعتبار سنجی)؛ 2) کیفیت داده های ورودی قابل جمع اوری (حجم، سرعت، نوع، صحت)؛ 3) قلمروی داده های فیدبک اعم از اثرات نه فقط خروجی ها.

نیرومندی. اینجا، تمرکز بر عملکرد مدل است- یک تابع داده های آموزشی (درستی یابی) و فیدبک (اعتبار سنجی) – و راهبردهای کاهش به ویژه بر خلاف مجهول های مجهول (دیتریچ، 2017).

سطح اتوماسیون. اینجا تاکید بر نقش های نسبی ماشین در برابر انسان در تصمیم و عملکرد است. ما اهمیت توضیحات (شفاف و پسین) را در اعتبار سنجی زمان اجرا، مورد تاکید قرار می دهیم، مثلا زمانیکه ممکن است انسان نیاز داشته باشد به خاطر یک معلوم مجهول بر ماشین برتری یابد. هم چنین مد نظر قرار دادن خطر اتکاء بیش از حد انسان به ماشین، به طور خاصی مهم می شود، ماشین باید به طور آشکار عدم قطعیت خود را منتقل نماید.

مالکیت. در این صورت، تاکید بر پتانسیل برای برونسپاری عناصر کلیدی سیستم تصمیم گیری بخش عمومی است، بویژه داده ها (ورودی، آموزش، فیدبک) و مدل، مثلا قرارداد با یک فروشنده هوش مصنوعی برای ساخت و مدیریت سیستم، ترک تصمیم و عملکرد داخلی[۱۲]. موافقین و مخالفین باید سنجیده شوند؛ عوامل، شامل یک کمبود تخصص هوش هوش مصنوعی/ یادگیری ماشین برای انجام آن به طور کلی در داخل و مدیریت داده های مناسب با برونسپاری در برابر حفظ ارزش در داده ها است.

ارزیابی و "آنچه درست عمل می کند". اینجا، تاکید بر قضاوت، تصمیمات، عملکردها و فیدبک است(من جمله اثرات): کسب بهترین شیوه ها و "آنچه خوب عمل می کند". این امر، داده ها را برای استفاده از مدل در شروع بخش قبلی به عنوان یک فرایند تصمیم به جای فرا فرایند ارائه می دهد، یعنی استفاده از هوش مصنوعی/ یادگیری ماشین برای انتخاب کاربردهای نوید بخش هوش مصنوعی/ یادگیری ماشین.

هوش مصنوعی برای تصمیم گیری بخش عمومی : راهبردها

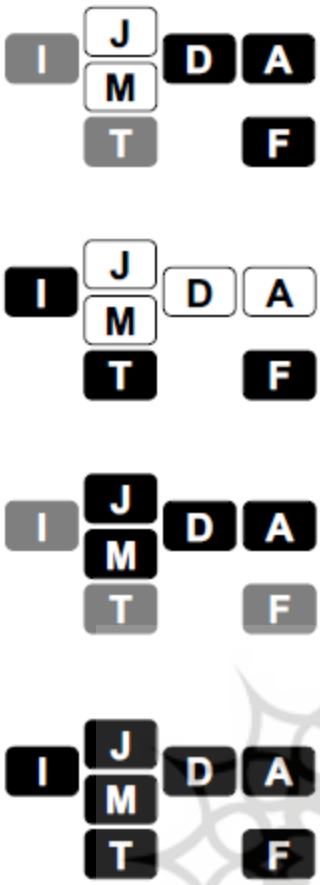
اکنون به ارزیابی چندین راهبرد برای پرداختن به تصمیم گیری بخش عمومی از طریق هوش مصنوعی/ یادگیری ماشین می پردازیم که در شکل 3 خلاصه شده اند. همانند قبل، در هر مورد، این مدل به صورت یک نسخه انتزاعی از شکل 1 با عناصر داده ای به رنگ خاکستری، عناصر عملیاتی به رنگ سفید و عناصر مورد تمرکز (داده ها یا عملیاتی) در هر صورت به رنگ مشکی نشان داده شده اند.

ماموریت گرا. این راهبرد به منظور به حداکثر رساندن کیفیت تصمیم گیری و به حداکثر رساندن ارزش از هوش مصنوعی/ یادگیری ماشین به ماموریت سازمان بخش عمومی است. از اینرو، تاکید بر پیامدهاست. تصمیم، عملکرد و فیدبک. توضیح، کاوش می کند، چرا چیزی که درست عمل کرده است، درست عمل کرده است و چرا درست عمل نکرده است.

داده گرا. اینجا، هدف، به حداکثر رساندن ارزش از دارایی های داده است و به این ترتیب، بر عناصر داده مدل، تمرکز دارد: آموزش، ورودی و فیدبک. این راهبرد، مدیریت و توسعه داده موثر را مورد تاکید قرار می دهد، چه به صورت داخلی، چه جمعی در سطح بخش یا به صورت مشارکت عمومی- خصوصی. هدف، افزایش فضای مجهول های معلوم تا بیشترین میزان ممکن و انتقال بار ارزیابی بیشتر برای درستی یابی است.

---

[12] In house



ماموریت گرا. بر پیامدها متمرکز است: داده های تصمیم، عملکرد و قیدیک.

داده گرا. بر عناصر داده متمرکز است: ورودی ها، آموزش و قیدیک

کار گرا. بر ترکیبی از قضاوت انسانی و مدل ML در رابطه با عناصر تصمیم و عملکرد.

ارزیابی گرا. به طور کل نگرانه بر حلقه کلی تمرکز دارد، به منظور کسب بهترین شیوه و چیزی است که به بهترین شکل عمل می کند.

کار گرا. این راهبرد، بر به حداکثر رساندن ارزش از همکاری انسان و ماشین تمرکز دارد یعنی تعامل بین مدل یادگیری ماشین و قضاوت انسان در تصمیم گیری و نقش های انسان / ماشین متعاقب در عملکرد. بنابراین، این راهبرد، به دقت، ماهیت کار انجام شده توسط انسان ها و ماشین ها را در سازمان، بررسی می کند، با دیدگاهی برای توسعه هر دارایی به موثرترین شکل ممکن این کار صورت می گیرد. توضیح برای اعتبار سنجی، شفافیت و اعتماد، کلیدی است. راهبرد، فرصت هایی را برای نیرومندی سیستمیک بهبود یافته بدست می آورد (مثلا بر خلاف معلوم های مجهول و مجهول های مجهول) و رضایت شغلی بهتر اگر به خوبی انجام شده باشد (دارایی های انسان، بهتر می توانند این تقاضاهای خدمت را به عهده گیرند).

ارزیابی گرا. این یک راهبرد کل نگرانه است: حلقه کل را با یک نیت کسب بهترین شیوه، در نظر می گیرد: "چیزی که خوب عمل می کند" اعم از بهترین شیوه در ارزیابی و توضیح. هدف، خلق پایگاه شواهد برای مداخلات هوش مصنوعی است.

بحث و نتیجه گیری

در ارزیابی سیستم مهم است که بین خروجی سیستم و تأثیرات سیستم تفاوت قائل شویم. در حالی که معیارهای متداول یادگیری ماشین مانند ماتریس های دقت و ابهام (درست / نادرست ،مثبت/ منفی) در مورد عملکرد یک مدل ML دارای اطلاعات مفیدی هستند ، این سوال که آیا سیستم تأثیر "درستی" بر سازمان دارد یا خیر ، یک سوال مجزا است. کل حلقه باید از منظر ادغام سازمانی و طراحی فرآیند مورد توجه قرار گیرد. به ندرت اتفاق می افتد که یک سیستم هوش مصنوعی به سادگی در فرآیندها / سازمان ها / سیستم های موجود "ادغام شود" و در این صورت انتظار می رود تأثیر مطلوبی داشته باشد. علاوه بر

این ، مدل تصمیم گیری انسان + ماشین در بالا فرض می کند که عنصر انسانی تصمیم گیری در سازمان هدف به خوبی مشخص شده است - اغلب ، در سازمانهای بزرگ ، تصمیم گیری توزیع می شود ، بنابراین درک تأثیر وارد کردن یک یا چند عنصر هوش مصنوعی به چنین سیستم جمعی بسیار چالش برانگیز می شود.

نیاز برای توضیح و ارزیابی باید برای اندازه گیری اثر، مناسب باشد. شفافیت، همیشه یک مساله مهم محسوب می شود. مثلا، در بسیاری از مداخلات پزشکی، مشخص نیست چطور پویایی در سطح مولکولی یا سیستم ها، درست عمل می کند (به طور خاص برای داروهای خاص)، اما با این وجود، دارای شواهد قوی برای نشان دادن این امر است که آنها به خوبی عمل می کنند. توضیحات پسین قانونی، ممکن است این مبنای شواهد را استنباط نماید. ساخت نوع مبنای شواهد مورد بحث در بخش قبلی می تواند یک اساس مشابه را برای تصمیم گیری بخش عمومی ارائه دهد که از هوش مصنوعی استفاده می کند: کارورزان می توانند با یگان های مورد اعتمادی کار کنند که از مبنای شواهد برای توصیه سیستم های هوش مصنوعی، ابزارها یا رویکردها استفاده می نمایند. با این حال، عنصر فیدبک مدل ما، باز هم مهم است: یک مداخله "اثبات شده" ممکن است، در یک زمینه نوین به خاطر عوامل ناشناخته،با شکست مواجه شود (به طور خاص، مجهول های معلوم و مجهول های مجهول). فیدبک، مدل هوش مصنوعی و مبنای شواهد را بهبود می بخشد. یادگیری حلقه واحد، دوگانه و سه گانه در اینجا، مهم محسوب می شود (مولگان، 2017). بار دیگر، این بحث به تمایز مهمی بین "درست سازی سیستم" (دقت بالا و ...) و "ساخت سیستم صحیح" (اثر بالا) می رسد. "زمستان هوش مصنوعی" که در اواخر دهه 1980شروع شد، به خاطر شکست این فناوری برای تامین انتظارات کاربر با وجود اغلب عملکردهای فنی بالا بود. فضای کاربردهای موثر با دوام، از آنچه توسعه دهندگان و سرمایه گذاران، امید داشتند، بسیار کوچکتر بود. یک مبنای شواهد قوی برای اثر هوش مصنوعی، برای به چالش کشیدن هیپ[۱۳] لازم است که یکبار دیگر پیشرفت های فنی در این حوزه را در بر می گیرد.

منابع و مراجع

Bratko, I. 1997. Machine learning: Between accuracy and interpretability. In Della, R. G.; Lenz, H.; and R., K., eds., Learning, Networks and Statistics (International Centre for Mechanical Sciences (Courses and Lectures), vol 382), volume 382, 163–177. Springer.

Cummings, M. 2014. Man versus machine or man + machine? IEEE Intelligent Systems 29(5):62–69.

Goodman, B., and Flaxman,

S. 2016. European Union regulations on algorithmic decision-making and a “right to explanation”. In 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), 26–30.

Lipton, Z. C. 2016. The mythos of model interpretability. In 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), 96–100.

Terveen, L. 1995. Overview of human-computer collaboration. Knowledge-Based Systems 8(2):67–81.

[13] hype