یادگیری عمیق و مشکل جعبه سیاه در هوش مصنوعی

مرحله دوم، یادگیری عمیق بود. یک مشکلی هم که الآن هست، مشکل جعبه سیاه است. الآن خیلی مهم است و دارند خیلی روی آن کار می‌کنند؛ «Black-box problem». این جعبه سیاه، شاهد عرض من است که تفلسف و ایده فلسفی، پشت کار پیشرفت اینها نیست.

الآن می‌خواهند کار شبکه عصبی را و نورون‌های عصبی را شبیه‌سازی کنند و پیاده کنند، این فلسفه است؟! من به گمانم اصلاً فلسفه نیست. در مهندسی نرم‌افزار می‌بینند که شبکه‌های عصبی رفتارشان به چه صورت است تا همان را پیاده کنند. اینجا اصلاً فلسفه نیست. او اصلا کاری با فلسفه ندارد. آنجا یک رفتاری را می‌خواهد از نورون‌ها ببیند و آن را پیاده کند. مرحله اولش ساده بود؛ نورون‌های ساده‌ای که تنها یک لایه دارد: ورودی و لایه پردازش و خروجی. به این، ساده‌ترین پردازش شبکه سلول عصبی می‌گویند. ورودی و یک لایه پردازش و خروجی.

شاگرد۳: این متوقف بر این است که معرفت را یک امر مجرد بدانیم.

استاد: اصلاً متوقف بر اینها نیست. مثال بزنم؛ مثلاً انانوتکنولوژی که الآن کار می‌کنند، چقدر مبتنی بر تفلسف است؟! ذهنمان را دور نبریم. فیزیک خیلی نزدیک به تفلسف است. از علومی که خیلی دور از تفلسف است -نمی‌گویم نیست - شیمی است. رواله کارهای شیمی و تحقیقاتی که دارند، تفلسف ندارد؛ می‌خواهم غلبه را بگویم. الآن انانوتکنولوژی مرتب دارد کار می‌کند - ذرات را مهندسی می‌کنند و مواد با خواص مختلف پدید می‌آید - فلسفه این چیست؟ شما بگویید.

شاگرد۳: بله، ممکن است از فلسفه دور باشد، اما باید ببینیم معرفت چطور است. چون می‌خواهد بشناسد.

استاد: ما کاری به معرفت نداریم. ما می‌خواهیم بگوییم یک شبکه نورونی با پیامدهای شیمیایی و الکتریکی چکار می‌کند. یک نورون عصبی، ورودی دارد و خروجی دارد. چکار می‌کند؟ شما می‌گویید معرفت چیست، درحالی‌که کاری به معرفت ندارد. دارد می‌گوید این عملکرد یک نورون است. بعداً عرض می‌کنم که مشکل جعبه سیاه چیست.

شاگرد۳: پیش‌فرض آنها این است که آگاهی از همین نورون‌ها به وجود می‌آید

استاد: نه، شما به قدم‌های آخر نروید. اگر من توضیح آن را عرض کنم، می‌بینید که آنچه که می‌گویید، چه آخر کار وحشتناکی است؛ برای خودش یک چیزی است. ابتدا همین است: او می‌خواهد کار یک نورون را شبیه‌سازی کند. نسل دوم این است.

شاگرد۲: این قدم اولش بود.

استاد: این قدم ساده اول بود. «Deep Learning» یا یادگیری عمیق است. آموزش، مرتبه بعدی است که عرض می‌کنم. یادگیری عمیق چیست؟ لایه‌ها، متعدد می‌شود. یعنی وقتی ورودی وارد می‌شود، مثلاً در ده لایه، پردازش می‌آید تا خروجی دربیاید. اینجا وقتی یادگیری عمیق پیش آمد، یعنی الآن مهندسین با ابزار شبیه‌سازی شبکه عصبی و نورون‌های مغزی، پردازشی عمیق درست کرده‌اند. یعنی ورودی‌ای که به آن می‌دادند، چند لایه پردازش می‌شد و بعد، خروجی می‌داد. اینجا بود که مشکل جعبه سیاه پیش آمد. یعنی مکانیسم اینکه چرا جواب داد، برای آن‌ها مخفی است. جعبه سیاه، یعنی نمی‌دانند که چه گذشت. جعبه شیشه‌ای و سفید و این‌ها نیست. الآن این را جلو بروی، تا بعد به آگاهی برسیم. این الآن برایشان خیلی ناخوشایند است. آقا هم می‌گفت یک جواب‌های با اطمینانی می‌دهد، دیدم خودشان نمی‌توانند برگردند و ببینند چرا این کارها را می‌کند. مهندس آمده در نسل سوم و بعد، نسل چهارم «Deep Learning» کرده، کار خودش را انجام می‌دهد و شروع می‌کند به او داده می‌دهد. آن هم حسابی جواب می‌دهد؛ یک جاهایی جواب‌های خوبی می‌دهد و یک جاهایی هم با اطمینان، جواب‌های غلط می‌دهد که از خودش درآورده است. می‌گویند توهم زده است. خود مهندسین می‌گویند ماشین توهم زده است، چرا؟ نمی‌دانند.

شاگرد۲: جهتش را نمی‌دانستند

استاد: اصلاً نمی‌توانند برگردند و ببینند چکار کرده‌اند. یعنی در آن درخت تصمیم‌گیری که یکی را انتخاب می‌کند، نمی‌دانند چرا این را انتخاب کرد.

شاگرد۲: این در گام دوم بود؟

استاد: بله، در یادگیری عمیق، مشکل جعبه سیاه پیش آمد.

شاگرد۲: این الآن همان چیزی است که در هوش مصنوعی‌ها سر و کار داریم؟

استاد: بله.

شاگرد۲: اینکه جواب غلط می‌دهد، همین است؟

استاد: بله، با اطمینان هم جواب غلط می‌دهد و خود مهندسین هم نمی‌توانند آن را تصحیح کنند. در هوش مصنوعی نمادین، برمی‌گردد و درست می‌کند و می‌گوید اینجا اشتباه شد. اما اینجا چون طبق اتصالات شبکه عصبی است، آن هم عمیق، یعنی ما نمی‌دانیم در لایه‌های مخفی پردازش چکار کرده که این خروجی را داده، لذا نمی‌توانیم برگردیم و بگوییم آن را درست کن. اینجا خیلی زحمت کشیده‌اند که چکار کنیم. به این، هوش مصنوعی توضیح‌ناپذیر می‌گویند. یعنی نمی‌تواند توضیح داد که چکار کرده است.

الآن دارند کار بسیار مهمی انجام می‌دهند؛ «xAI» می‌گویند، یعنی توضیح‌پذیر؛ می‌خواهند هوش مصنوعی توضیح‌ناپذیر را توضیح‌پذیر کنند.

شاگرد۴: نسل سوم چه می‌شود؟

استاد: نسل سوم همین آموزش عمیق است.

شاگرد۴: منظورتان از نسل سوم همین لایه دوم است؟

استاد: نسل‌ها را که به کار می‌برند اصطلاحات مختلفی دارد. فقط باید محتوا را دید. البته اندازه‌ای که من مطالعه کرده‌ام.

شاگرد۴: اینکه با فاصله کوتاهی بعد از نسل دوم، نسل سوم آمد... .

استاد: نسل سوم نه، گفتم مرتبه دوم نسل دوم. مبنای نسل دوم، اتصال‌گرائی و شبکه‌های عصبی است. وقتی در فضای شبکه‌های عصبی وارد می‌شوید، دیگر سر و کار شما با معنا نیست. این‌ها را عرض می‌کنم چون قبلاً گفتیم که یک شبکه معنا درست می‌کنند. شبکه معنایی که من عرض کردم، دو جور است. الآن یک شبکه معنایی هست که روی همان مبنای قاعده‌محور است که آن، یک جور شبکه درست می‌کند و قبلاً توضیح آن را عرض کردم. طبق حرف‌هایی که الآن هست، شبکه معنایی، طور دیگری به پا می‌شود.

شاگرد۲: الآن این گام دوم را پیش رفته‌اند؟ متوجه شده‌اند که چرا جواب غلط می‌دهد؟

استاد: نه، متوجه نمی‌شوند. اصلاً نخواه‌ند شد. یعنی مشکلی است که نمی‌توانند کاری کنند. او خودش پردازش انجام می‌دهد و آن‌ها نمی‌توانند بفهمند چکار کرده است. فقط می‌توانند مرحله به مرحله تفکیک کنند و آن را توضیح‌پذیر کنند؛ «xAI» کنند.

بازبینی #2
ایجاد شده 9 دسامبر 2025 15:05:25 توسط ... .
به روزرسانی شده 9 فوریه 2026 16:17:24 توسط ... .