第 5 章CHAPTER 5價值矩陣——人工智能的準確性是胡說八道。以下是使用者體驗必須採取的措施Value Matrix—AI Accuracy Is Bullshit. Here ’ s What UX Must Do About It

數據科學社區中最保守的秘密是什麼？人工智慧的準確性在現實世界中毫無意義。本章為使用者體驗設計師提供了一種實用的準確性替代方案：一種以使用者體驗和業務為中心的方式來優化您的人工智慧解決方案，以「人類」價值觀來「思考」。

What ’ s the best-kept secret in the data science community? AI accuracy is meaningless in the real world. This chapter gives UX designers a practical alternative to accuracy: a UX- and business-centric way to optimize your AI solution to “ think ” in terms of “ human ” values.

你準備好了嗎？

Are you ready?

大秘密

The Big Secret

多年來，資料科學世界一直以準確度、精確度和召回率等資料科學指標為運作 （請參閱本章後面的側邊欄「精確度、召回率和準確度」，以取得這些術語的詳細說明）。像 Kaggle （1） 這樣的數據科學競賽僅根據準確性等單一指標來確定獲勝者。價值矩陣由 Arijit Sengupta 專門針對 AI 的實際應用開發 （2-5）。

For many years, the data science world operated on data science metrics like accuracy, precision, and recall (see the sidebar “ Precision, Recall, and Accuracy, ” later in this chapter, for a detailed explanation of these terms). Data science competitions like Kaggle (1) determine winners exclusively on a single metric like accuracy. The Value Matrix was developed by Arijit Sengupta specifically for real-world applications of AI (2 – 5).

便條

NOTE

最大的秘密是，數據科學指標對人工智能的實際應用意義不大。當談到準確性等指標時，它們通常是完整且完全廢話。

The big secret is that data science metrics mean little to real-world applications of AI. When it comes to metrics like accuracy, they are most often complete and utter bullshit.

讓我們舉一個簡單的用例：汽車保養。想像一個虛構的汽車製造商，“ Pascal Motors ”。（為什麼是布萊斯・帕斯卡？好吧，由於特斯拉和沃爾特汽車都被收購了，我們決定深入挖掘我們的歐洲老科學家寶庫。Pascal Motors 製造的汽車配備了車載人工智能，每當汽車需要進行定期維護時，該人工智能就會發出特殊警報。假設在運行一年內，一輛典型的汽車有 100 個總潛力和 20 個實際問題。我們還假設成功識別和預防問題的好處是 1,000 美元（例如，預防性維修：在零件發生故障並可能導致事故或汽車在高速公路中間拋錨的成本之前更換零件），調查潛在問題的成本為 100 美元（例如一名機械師檢查問題一小時的成本）。

Let ’ s take a simple use case: car maintenance. Imagine a fictional car manufacturer, “ Pascal Motors. ” (Why Blaise Pascal? Well, since Tesla and Volta Motors are both taken, we decided to dig deep into our reservoir of old European science dudes.) Pascal Motors makes cars with an onboard AI that sends out a special alert whenever a car needs to come in for scheduled maintenance. Let ’ s say in a year of operation, a typical car has 100 total potential and 20 actual problems. Let ’ s also assume that the benefit of identifying and preventing a problem successfully is $1,000 (e.g., preventive repair: replacing a part before it fails and potentially causes an accident or the cost of the car breaking down in the middle of the freeway), and the cost of investigating a potential problem is $100 (as in the cost of one mechanic checking out the problem for one hour).

Pascal Motors 工程師可以使用三種不同的 AI 模型：保守型、平衡型和激進型。這些模型具有以下資料科學指標，如圖 5.1 所示。

Pascal Motors engineers have access to three different AI models: Conservative, Balanced, and Aggressive. These models feature the following data science metrics, shown in Figure 5.1.

AI Model	Conservative	Balanced	Aggressive
Alerts Sent	10	30	80
Problems Found	9	15	19
Precision	**90%**	**50%**	**24%**
Recall	**45%**	**75%**	**95%**
Accuracy	**88%**	**80%**	**38%**

圖 5.1 基於資料科學指標的 AI 模型選擇：精確度、召回率和準確性

Figure 5.1 AI model selection based on data science metrics: precision, recall, and accuracy

您認為哪種 AI 模型最好？

Which AI model do you think is the best one?

大多數人會選擇保守派人工智慧，因為誰不想要一個既準確又精確的人工智慧呢？

Most people would pick the Conservative AI because who does not want an AI that is both accurate and precise?

看看圖 5.2。現在怎麼樣？

Take a look at Figure 5.2. How about now?

Model	Conservative	Balanced	Aggressive
Alerts Sent	10	30	80
Problems Found	9	15	19
Precision	**90%**	**50%**	**24%**
Recall	**45%**	**75%**	**95%**
Accuracy	**88%**	**80%**	**38%**
TP (+$1,000)	9 x $1,000	15 x $1,000	19 x $1,000
TN (+$100)	79 x $100	65 x $100	19 x $100
FN (-$1,000)	11 x -$1,000	-5 x $1,000	1 x $1,000
FP (-$100)	1 x -$100	-15 x $100	-61 x $100
Revenue	**$5,800**	**$15,000**	**$13,800**

圖 5.2 基於真實世界結果的 AI 模型選擇，假設 TP（真陽性）為 1,000 美元，TN（真陰性）為 100 美元

Figure 5.2 AI model selection based on real-world outcomes, assuming TP (true positive) of $1,000 and TN (true negative) of $100

如果您僅根據最佳數據科學指標選擇最佳 AI，那麼您的選擇將是錯誤的。

If you picked the best AI based solely on the best data science metrics, your choice would be wrong.

相較之下，如果我們透過考慮現實世界結果的實際成本和收益來優化收入，那麼正確的答案實際上是平衡型人工智慧（第 2 欄），它產生的收入比保守型（準確）人工智慧高出 158% 以上。

In contrast, if we instead optimize for revenue by taking into account the actual cost and benefit of real-world outcomes, the right answer is actually the Balanced AI (column 2), which produces over 158 percent more revenue than the Conservative (Accurate) AI.

便條

NOTE

僅根據資料科學指標進行最佳化的人工智慧幾乎總是表現不佳，因為人工智慧考慮了現實世界結果的成本和收益。

AI optimized on data science metrics alone will almost always underperform AI that considers the costs and benefits of real-world outcomes.

就是這樣。這是大秘密。

That ’ s it. That ’ s the big secret.

混淆矩陣：準確的人工智慧怎麼會出錯？

Confusion Matrix: How Can Accurate AI Be Wrong?

在這一點上，您可能會感到困惑。準確的人工智慧到底有多大錯誤？精準的AI不就是目標嗎？為了回答這個問題，我們需要深入研究一個簡單的公式來計算準確性，但我保證它會

很快——我們會盡可能簡單地進行討論，所以即使你對數學課沒有太多愉快的回憶，請繼續閱讀——我保證這是值得的。

At this point, you might be confused. Just how exactly can accurate AI be wrong? Isn’t accurate AI the goal? To answer this question, we need to dig into a simple formula of how accuracy is calculated, but I promise it will be quick—and we’ll make the discussion as simple as we can, so even if you don’t have a lot of pleasant memories of your math classes, please read on—I promise it will be worth it.

現在，要理解「準確性」的含義，我們必須看一個稱為混淆矩陣的簡單表格。混淆矩陣遠非令人困惑，它實際上非常簡單：它只是一個表格，我們收集模型預測的計數並將其與實際結果進行比較。

Now, to understand what “accuracy” means, we have to look at a simple table known as a Confusion Matrix. Far from being confusing, the Confusion Matrix is actually pretty straightforward: It is simply a table where we collect the counts predicted by the model and compare them against the actual outcomes.

每次 Pascal Motors AI 查看汽車的 100 個潛在問題時，它都可以決定是否發出警報。當人工智慧決定發出警報時，這是一個積極的因素。如果人工智慧決定忽略感測器讀數，則表示負面。如果我們假設一年內有 100 個潛在事件，那麼我們的 AI 總共有 100 個決策點，它可能會決定是否發送警報。

Every time Pascal Motors AI looks at 100 potential problems with a car, it can decide whether or not to send out an alert. When the AI decides to send out an alert, that’s a positive. If the AI decides to ignore a sensor reading, it’s a negative. If we assume there are 100 potential events in one year, our AI has a total of 100 decision points where it can potentially decide whether to send out an alert.

現在，我們的人工智慧實際上並不能確定汽車有問題。它必須依賴各種傳感器的讀數，例如發動機油雜質、振動、奇怪的聲音等。因此，有時它可能會猜錯，並在不應該發出警報時發出警報，這稱為誤報。例如，如果一輛汽車在寒冷的早晨聽起來很奇怪，而我們的 AI 確定有問題，但機械師沒有發現，則可能會出現誤報。

Now our AI does not actually “know” for sure there is a problem with a car. It has to rely on the readings of various sensors, like, say, engine oil impurities, vibration, weird sounds, etc. So sometimes it might guess wrong and send out an alert when it should not have—that’s called a false positive. A false positive might occur, for example, if a car just sounded weird on a cold morning and our AI decided there was a problem but mechanics found none.

相反，人工智慧可能會錯過可能出現問題的情況，並在汽車實際上即將發生嚴重故障時判斷汽車運作正常。在這種情況下，人工智慧會錯誤地錯過發送警報，從而產生誤報。

Conversely, the AI might miss a condition that might be a problem and decide the car is operating properly when it is actually about to have a serious breakdown. In this case, the AI will erroneously miss sending out an alert, creating a false negative.

因此，對於給定年份的 100
個潛在決策點中的每一個，人工智慧可能會得出四種可能的結果：

Thus for every one of the 100 potential decision points in a given year, the AI might come up with four possible outcomes:

- 真陰性（TN）：汽車沒有問題，AI不發送警報。True Negative (TN): There is no problem with the car, and AI does not send an alert.
- 假陰性（FN）：其實有問題，但AI沒有告訴我們。False Negative (FN): There is actually a problem, but AI does not tell us.
- 真陽性 （TP）：汽車有問題，AI 正確發送警報。True Positive (TP): There is a problem with a car, and AI sends the alert correctly.
- 誤報 （FP）：汽車運行正常，但 AI 發出警報。False Positive (FP): The car is operating normally, but AI sends out an alert.

混淆矩陣只是特定模型產生的每個結果的計數。這個矩陣是一個有用的工具，因為它允許我們透過比較不同人工智慧模型產生的各種結果的不同計數來了解它們的表現。

The Confusion Matrix is simply a count of each of the outcomes a particular model generates. This matrix is a useful tool because it allows us to see how different AI models perform by comparing the different counts of various outcomes they generate.

很簡單，不是嗎？

Simple, no?

因此，對於我們之前查看的保守（高度準確）AI 模型，混淆矩陣如圖 5.3 所示。

Thus, for the Conservative (highly accurate) AI model we looked at earlier, the Confusion Matrix is shown in Figure 5.3.

	Predicted: NO	**Predicted: YES**	
Actual: NO	TN 79	FP 1	80
Actual: YES	FN 11	TP 9	20
	90	10	**100**

圖 5.3 保守 AI 模型的混淆矩陣

Figure 5.3 The Confusion Matrix for the Conservative AI model

要閱讀混淆矩陣，請先繞過表格的外部。如前所述，我們總共有 100 次測量和 20 張實際贊成票，這意味著 80 張實際反對票。保守的AI發出了10個警報，預測了90次 “ 沒有問題 ” 。在保守派AI發出的10個警報中，它正確預測了9個問題（真陽性），並做出了1個錯誤的警報預測（誤報）。

To read the Confusion Matrix, start by going around the outside of the table. As discussed earlier, we had a total of 100 measurements and 20 actual yes votes, which means 80 actual no ’ s. Conservative AI sent out 10 alerts and predicted that there was “ no problem ” 90 times. From the 10 alerts the Conservative AI sent out, it correctly predicted 9 problems (true positive) and made 1 incorrect alert prediction (false positive).

從這個方程計算精度非常簡單。我們將正確預測的總數除以預測總數，然後以百分比表示。

To compute accuracy from this equation is pretty straightforward. We take the total number of correct predictions and divide that by the number of total predictions and express it as a percentage.

Accuracy = Correct Predictions/Total Predictions * 100%

舉個簡單的例子，如果你總共拋硬幣 100 次，並預測每次拋硬幣的“ 正面 ”，那麼你大約會正確 50 次，所以平均準確率為 50%。

To use a simple example, if you tossed a coin a total of 100 times and predicted “ heads ” on every coin toss, you would be correct about 50 times, so you ’ d be 50 percent accurate, on average.

以 Pascal Motors 為例，總共 20 次測量中有 100 個實際問題。因此，我們的保守 AI 在總共 100 個預測中總共做出了 88 個正確的預測（79 個真陰性 + 9 個真陽性），因此其準確性如下：

In the case of Pascal Motors, there were 20 actual problems from 100 total measurements. Thus our Conservative AI made a total of 88 correct predictions (79 true negative + 9 true positive) out of a grand total of 100 predictions, so its accuracy is as follows:

Accuracy = (79 + 9)/100 * 100% = 88%

現在 88% 的準確度很高！然而，對我們來說不幸的是，該模型遺漏了 11 個可能的問題中的 20 個。事實上，保守的 AI 模型對我們來說其實一點用不上——這個 AI 模型發現的問題還不到一半！

Now 88 percent is a great accuracy! Unfortunately for us, however, the model missed 11 out of 20 possible problems. In fact, the Conservative AI model is actually less than useless for us—this AI model found less than half the problems!

準確的人工智慧怎麼變得如此無用？到目前為止，答案應該不會讓您感到驚訝。

How does an Accurate AI become so useless? By now, the answer should not surprise you.

AI trained on accuracy is often too timid: It tries too hard not to be wrong, and so it " leaves money on the table " by not taking enough chances to send out an alert.

相反，在另一個極端，

Conversely, on the other extreme,

AI that is trained on recall tries to account for every possible positive—which is often too aggressive for real-world use.

例如，為了找出 20 個問題中的 19 個，我們的 Aggressive AI 模型發出了 80 個警報！你真的能想像一個顧客每 365/80 = 4.5 天跑到一家汽車商店嗎？

For instance, in an effort to locate 19 out of 20 problems, our Aggressive AI model sent out 80 alerts! Can you really imagine a customer running to a car shop every 365/80 = 4.5 days?

實際應用的最佳 AI 選擇實際上是平衡 AI;儘管它在任何特定的數據科學指標上並不出色，但它產生了最高的投資回報率，為 15,000 美元，比準確模型高出 158% 以上。

The best AI choice for a real-world application is actually Balanced AI; although it does not excel in any particular data science metric, it produces the highest ROI, $15,000—which is more than 158 percent higher than the Accurate model.

And in the real world, the ROI is the only metric that actually matters.

好的，希望您現在確信，僅使用準確性和召回率等數據科學指標並不能為您提供適用於實際應用的最佳 AI。開發正確的人工智慧需要一個不同的工具：價值矩陣。

Okay, hopefully, you are now convinced that using just data science metrics like accuracy and recall are not going to give you the best AI for real-world applications. What you need to develop the right AI is a different tool: the Value Matrix.

價值矩陣：現實世界的人工智慧工具

Value Matrix: The AI Tool for the Real World

價值矩陣由 Arijit Sengupta 專門針對 AI 的實際應用開發（2-5）。價值矩陣是對傳統混淆矩陣的簡單調整。顧名思義，在價值矩陣中，UX 設計師或產品經理以美元形式記錄每個結果的價值，然後將該值乘以混淆矩陣中每個結果的計數，讓我們清楚地了解整體 AI 模型的投資回報率。

The Value Matrix was developed by Arijit Sengupta specifically for real-world applications of AI (2 – 5). The Value Matrix is a simple tweak to the traditional Confusion Matrix. As the name implies, in the Value Matrix, the UX designer or product manager records the value of each outcome in dollar terms, then multiplies this value by the count of each outcome in the Confusion Matrix, giving us a clear reading of the overall AI model ROI.

例如，使用保守的 AI 模型的混淆矩陣，假設成功識別和預防問題的收益為 1,000 美元，調查潛在問題的成本為 100 美元，則相應的價值矩陣將如圖 5.4 所示。

For instance, using the Conservative AI model ’ s Confusion Matrix and assuming that the benefit of identifying and preventing a problem successfully is $1,000 and the cost of investigating a potential problem is $100, the corresponding Value Matrix would look like the one shown in Figure 5.4.

正確的 AI 猜測是正面的（收益），錯誤的猜測是負面的（成本）。例如，在我們目前的一組假設中，在沒有問題的情況下將客戶送到維修店可能會花費公司 100 美元（例如，－100 美元）。相反，正確識別沒有問題的結果可以節省 100 美元（例如，+100 美元）。正確識別問題（真陽性）可節省 1,000 美元（例如，+1,000 美元），而錯過問題則節省 1,000 美元（例如，－1,000 美元）。

Correct AI guesses are positives (benefit), and wrong guesses are negatives (cost). For example, in our current set of assumptions, sending a customer into a repair shop when there is no problem might cost the company $100 (e.g., − $100). Conversely, correctly identifying the outcome where there is no issue generates $100 of savings (e.g., +$100). Correctly identifying a problem (true positive) saves $1,000 (e.g., +$1,000), whereas missing the problem costs $1,000 (e.g., − $1,000).

	Predicted: NO	**Predicted: YES**	
Actual: NO	TN 79 +$100 +$7,900	FP 1 –$100 –$100	80
Actual: YES	FN 11 –$1,000 –$11,000	TP 9 +$1,000 +$9,000	20
	90	10	**100**

圖 5.4 保守（準確）AI 模型的混淆矩陣，假設 TP（真陽性）為 1,000 美元，TN（真陰性）為 100 美元

Figure 5.4 The Confusion Matrix for the Conservative (Accurate) AI model, assuming TP (true positive) of $1,000 and TN (true negative) of $100

便條

NOTE

雖然此特定範例的 TP/TN 和 FP/FN 結果值相同，但沒有規則總是如此。在其他使用案例中，每個結果的美元值可能不同。（換句話說，咳咳，您的里程可能會有所不同。

Although the values of TP/TN and FP/FN outcomes are the same for this particular example, there is no rule that it is always so. The dollar values for each outcome might be a different number in other use cases. (In other words, ahem, your mileage might differ.)

從本質上講，價值矩陣是一種工具，可以幫助您的團隊認識到每個預測結果都會產生金錢效應。價值矩陣非常強大，因為它使我們能夠評估部署不同人工智慧模型的實際結果。

Essentially, the Value Matrix is a tool that helps your team recognize that each predictive outcome produces a monetary effect. The Value Matrix is exceptionally powerful because it allows us to evaluate the real-world outcomes of deploying different AI models.

根據現實生活結果訓練人工智慧像人類一樣「思考」

Training AI on Real-Life Outcomes to “ Think ” Like a Human

到目前為止，很明顯，不同的價值假設會產生一個非常不同的價值矩陣。例如，如果我們的用例中誤報的成本更高（假設客戶每次進入商店時花費 800
美元），您會對具有最高準確性的保守 AI 模型感到非常滿意，該模型力求不會出錯。圖 5.5 顯示了這個新的價值矩陣細分。

By now, it should be evident that a different value assumption would produce a very different Value Matrix. For instance, if the cost of a false positive in our use case was higher—say it cost the customer $800 every time they came into the shop—you ’ d be pretty happy with a Conservative AI model with the highest accuracy, which seeks to not be wrong. Figure 5.5 shows this new Value Matrix breakdown.

Model	Conservative	Balanced	Aggressive
Alerts Sent	10	30	80
Problems Found	9	15	19
Precision	**90%**	**50%**	**24%**
Recall	**45%**	**75%**	**95%**
Accuracy	**88%**	**80%**	**38%**
TP (+$1,000)	9 x $1,000	15 x $1,000	19 x $1,000
TN (+$800)	79 x $800	65 x $800	19 x $800
FN (-$1,000)	11 x -$1,000	5 x -$1,000	1 x -$1,000
FP (-$800)	1 x -$800	15 x -$800	61 x -$800
Revenue	**$60,400**	**$50,000**	**-$15,600**

圖 5.5 基於真實世界結果的 AI 模型選擇，假設 TP 為 1,000 美元，TN 為 800 美元

Figure 5.5 AI model selection based on real-world outcomes, assuming TP of $1,000 and TN of $800

相較之下，如果真陽性的價值更大，假設我們每次 AI 能夠預測問題時都會節省 10,000 美元，我們會希望我們的 AI 針對每個可能的潛在問題發出警報，因此訓練在召回率上的 Aggressive AI 會更好，因為這樣的模型試圖捕捉每一個可能的真陽性（見圖 5.6）。

In contrast, if the value of true positive were greater, say we would save $10,000 each time AI was able to predict a problem, we would want our AI to send out an alert on every possible potential issue, so the Aggressive AI trained on recall will be better, because such a model seeks to capture every possible true positive (see Figure 5.6).

Model	Conservative	Balanced	Aggressive
Alerts Sent	10	30	80
Problems Found	9	15	19
Precision	**90%**	**50%**	**24%**
Recall	**45%**	**75%**	**95%**
Accuracy	**88%**	**80%**	**38%**
TP (+$10,000)	9 x $10,000	15 x $10,000	19 x $10,000
TN (+$100)	79 x $100	65 x $100	19 x $100
FN (-$10,000)	11 x -$10,000	-5 x $10,000	1 x $10,000
FP (-$100)	1 x -$100	-15 x $100	-61 x $100
Revenue	**-$12,200**	**$105,000**	**$175,800**

圖 5.6 基於真實世界結果的 AI 模型選擇，假設 TP 為 10,000 美元，TN 為 100 美元

Figure 5.6 AI model selection based on real-world outcomes, assuming TP of $10,000 and TN of $100

比較圖 5.5 和圖 5.6。請注意，如果我們調整結果的成本/收益值，在每種情況下，在相反目標（保守與激進）上訓練的人工智慧模型實際上會產生負投資報酬率！

Compare Figure 5.5 and Figure 5.6. Note that if we tweak the cost/benefit values of the outcomes, in each of these cases, the AI models trained on the opposite goals (conservative vs. aggressive) actually produce a negative ROI!

例如，如果真陽性的值為 10,000 美元，而真陰性的值為 100 美元：

For example, if the value of true positive was $10,000 and the true negative was $100:

Deploying our Conservative, highly accurate AI model will actually cost our company $12,200 vs. generating $105,000 and $175,800 revenue if we deploy the other two models!

您仍然認為像準確性這樣的純數據科學指標與現實世界有任何相關性嗎？

Do you still think pure data science metrics like Accuracy have any relevance in the real world?

再舉一個例子

One More Example

作為用戶體驗和產品領導者，我希望您能夠輕鬆看到這一重要原則的大量應用，以及非常詳細地了解特定用例的每種結果的成本和收益的重要性。使用者體驗研究和分析對於幫助人工智慧以更「人性化」的方式思考至關重要。在這本書中，我為您提供了做到這一點的工具。

As a UX and product leader, I hope you can easily see a great number of applications of this important principle and the importance of understanding the costs and benefits of each outcome for your specific use case in great detail. UX research and analysis are essential for helping AI think in more “ human ” terms. And in this book, I give you tools to do exactly that.

便條

NOTE

人工智慧不會問：「哪個事件最有可能成為問題？」相反，人工智慧應該問一個商業問題：「我如何實現收入最大化？」使用投資報酬率而不是科學指標來了解人工智慧對現實世界的影響，使我們能夠訓練人工智慧像人類一樣「思考」。

Instead of AI asking: “ Which event is most likely to be a problem? ” AI should instead be asking a business question: “ How do I maximize revenue? ” Understanding the AI ’ s impact in the real world using the ROI instead of the science metrics gives us a handle on training AI to “ think ” like a human.

我將給您留下一個最壯觀的例子，強調為人工智慧預測賦予現實世界價值的重要性。這個例子來自 Arijit Sengupta：

I am going to leave you with one of the most spectacular examples underscoring the importance of attaching real-world value to AI predictions. This example comes from Arijit Sengupta:

假設 TSA 有一個人工智慧可以預測某人是否是恐怖分子。如果這個 AI 在 100% 的情況下返回錯誤，那麼它將是一個高度準確的 AI，準確率為 99.9999%，因為絕大多數通過 TSA 檢查站旅行的人都不是恐怖分子。

Assume the TSA had an AI that was predicting whether someone was a terrorist. If this AI returned false 100 percent of the time, it would be a highly accurate AI, at 99.9999999999999999 percent accuracy, because a vast majority of people traveling through a TSA checkpoint are not terrorists.

這樣的模型將是超級準確的！而且（顯然）超級無用。

Such a model would be super-accurate! And also (obviously) super-useless.

另一方面，如果這個 TSA AI 還考慮了恐怖襲擊的影響（約 1 兆 （6））與將可疑人員拉到一邊進行二次檢查的成本（可能是 TSA 特工時間的 2 分鐘，所以如果他們每小時支付 1 美元，則為 30 美元），你可以看到一個非常不同的 TSA AI 模型將會出現。他們可能想要針對召回率進行最佳化，而不是針對準確性進行最佳化，例如讓模型更具侵略性。更具侵略性......事實上，TSA 可以拉出 999,999,999 人（或整個地球上 80 億人口......乘以 125 倍！並且仍然領先 1 美元。

On the other hand, if this TSA AI also considered the impact of a terrorist attack (about $1 trillion (6)) vs. the cost of pulling a suspicious person aside for a secondary inspection (maybe 2 minutes of a TSA agent ’ s time, so $1 if they are paid $30/hour), you can see that a very different TSA AI model would emerge. Instead of optimizing for accuracy, they might want to optimize for recall, such as making the model more aggressive. Much more aggressive … In fact, the TSA could pull aside 999,999,999 people (or the entire Earth ’ s human population of 8 billion people… Multiplied 125 times!) and still come out $1 ahead.

這就引出了一個更深層次的問題：

Which begs a deeper question:

為什麼 TSA 不對每位旅客進行二次檢查？你怎麼看？

Why doesn ’ t the TSA do a secondary inspection for every traveler? What do you think?

最後的想法：人類成本/收益的重要性

Final Thoughts: The Importance of Human Cost/Benefit

自然地，我可以聽到使用者體驗設計師的合唱聲：「但是人類呢？我們客戶的投資報酬率如何？」

Naturally, I can hear the chorus of UX designers shouting, “ But what about the humans? What about our customers ’ ROI? ”

當然，你們都是對的。除了業務成本/收益之外，您還應該認真考慮人力/客戶成本/收益。

You all would be correct, of course. In addition to the business cost/benefit, you should think long and hard about the human/customer cost/benefit.

TSA 不會僅僅因為人力成本太高而對每位旅客進行二次檢查。人們通常必須在檢查隊伍中花費一個小時或更長時間，並且必須在每次航班起飛前四到五個小時到達機場。檢查人員的成本將飆升。到處都會擁堵，給機場設施帶來壓力。總體而言，對航空旅行業和整體旅行的負面影響將達到數千億美元，投資回報率都值得懷疑。

TSA does not do secondary inspection for every traveler simply because the human costs would be too high. People would routinely have to spend an hour or more in inspection lines and would have to show up to the airport four to five hours ahead of every flight. Costs for inspection personnel would skyrocket. There would be congestion everywhere, putting a strain on airport facilities. Overall, the negative impact on the air travel industry and travel in general would be counted in hundreds of billions, all with dubious ROI.

這就是為什麼使用者體驗對於創建人工智慧解決方案如此重要。

That is why UX is so essential to creating AI solutions.

我們更大的觀點是：

Our larger point is that:

便條

NOTE

人工智慧太重要了，不能把它留給資料科學家。僅靠準確度、精確度和召回率等純粹的數據科學指標並不能創建可行的現實世界解決方案。每個現實世界的人工智慧解決方案都應該透過對其產生的業務和人類影響的更深入理解來調整。人工智慧無疑是我們共同的未來——沒有什麼可以改變這一點。了解如何使用這個令人難以置信的工具來造福人類，並確保負責人為人類和地球做正確的事情，是您作為用戶體驗設計師工作的一部分。

AI is just too important to leave it to data scientists. Pure data science metrics like accuracy, precision, and recall alone don ’ t create viable real-world solutions. Every real-world AI solution should be tempered by a deeper understanding of both the business and human impact it creates. AI is indisputably our collective future—nothing can change that. Understanding how to use this incredible tool for the benefit of humankind and ensuring people in charge do the right thing for humanity and the planet is part of your job as a UX designer.

設計練習：創建自己的價值矩陣

Design Exercise: Create Your Own Value Matrix

現在輪到您為您的用例創建價值矩陣模型了。就像上一章的數位孿生一樣，價值矩陣也是一個小組活動。在「現實世界」中，您將與您的團隊一起執行此操作。但是，不要僅僅因為您是獨自一人而跳過此練習——這是解鎖 UX 以實現 AI 專案成功並從根本上增加您對團隊的價值的關鍵。若要在短短 10 分鐘內為您的使用案例建立價值矩陣，請回答下列一組簡單的問題：

Now it ’ s your turn to create a Value Matrix model for your use case. Just like the digital twin in the previous chapter, the Value Matrix is also a group activity. In the “ real world ” you would be doing this with your team. However, do not skip this exercise just because you are solo—it is the key to unlocking UX for AI project success and radically increasing your value to your team. To create a Value Matrix for your use case in just 10 minutes, answer the following simple set of questions:

- 真陽性有什麼好處？What is the benefit of a true positive?
- 真陰片有什麼好處？What is the benefit of a true negative?
- 誤報的代價是多少？What is the cost of a false positive?
- 假陰性的代價是多少？What is the cost of a false negative?
- AI能錯多少次，還能領先？How many times can the AI be wrong and still come out ahead?
- 您的專案需要保守（準確）或激進（高召回率）的 AI 模型嗎？Do you need a Conservative (accurate) or Aggressive (high recall) AI model for your project?

回想一下，TSA 不會對每位旅客進行二次檢查。照：

Recall that the TSA does not do a secondary inspection for every traveler. Reflect:

- 業務投資報酬率中蘊含的人力成本是什麼？What are the human costs embedded in the business ROI?
- 人類會因人工智慧的決定而感到過度不便嗎？Will humans be unduly inconvenienced by AI ’ s decisions?

- AI 模型的決策將對使用者體驗和長期客戶忠誠度產生什麼影響？What will be the impact of the AI model ’ s decisions on the UX and long-term customer loyalty?
- 人工智慧的決定是否合乎道德？在這種情況下，人類會如何決定？Is the AI decision ethical? How would a human decide in this situation?
- 這種人工智慧模型如何可能被濫用？How can this AI model potentially be misused?

如果您需要靈感，請看以下範例。在完成自己的設計練習之前，不要繼續下一章。

If you need inspiration, look at the following example. Do not proceed to the next chapter until you complete your own design exercise.

設計練習範例：生命時鐘值矩陣

Design Exercise Example: Life Clock Value Matrix

回想一下我們第 3 章中的 “ 生命時鐘 ” 故事板和第 4 章中的數字孿生示例。這是我們人工智慧驅動產品的價值矩陣。在本練習中，我們將特別關注系統中從手機圖像中預測食物類型的部分。首先，讓我們定義混淆矩陣。回想一下，這是一個簡單的表格，列出了 AI 猜測的四種可能結果，即 TP/TN/FP/FN：

Recall our “ Life Clock ” storyboard from Chapter 3 and our digital twin example in Chapter 4. Here ’ s the Value Matrix for our AI-driven product. In this exercise, we are going to specifically focus on the part of the system that predicts the type of food from the cell phone image. First, let ’ s define the Confusion Matrix. Recall that this is a simple table that lists the four possible outcomes of an AI guess as TP/TN/FP/FN:

真陽性 （TP）：AI 正確猜測並返回食物類型（ “ 這是 1 杯煮熟的燕麥片。-2 膽固醇。+7 無聊。

True Positive (TP): AI correctly guessed correctly and returned the food type (“ This is 1 cup of cooked oatmeal. – 2 Cholesterol. +7 to Boredom. ”)

真陰性 （TN） ：AI 正確猜測圖片不是食物並傳回錯誤。（ “ 這似乎不是一張食物的照片。您是否不小心上傳了迪士尼樂園雷山鐵路上家人的動作鏡頭，而不是超頹廢的貝奈特餅和薄荷朱利酒？-10 的耐力。

True Negative (TN): AI correctly guessed that the picture is not food and returned an error. (“ This does not appear to be a picture of food. Did you accidentally upload your family ’ s action shot from the Thunder Mountain Railroad in Disneyland instead of your ultra-decadent Beignets and Mint Julep splurge? −10 to stamina. ”)

誤報 （FP）：人工智慧錯誤地猜測了食物類型。（“ 這是意大利辣香腸披薩 ”，而您上傳了一張在花椰菜披薩皮上放養的純素豆腐甜菜香腸和腰果奶酪的照片。

False Positive (FP): AI incorrectly guessed the food type. (“ This is a pepperoni pizza, ” whereas you uploaded a picture of a free-range vegan tofu beet sausage with cashew cheese on a cauliflower pizza crust.)

假陰性 （FN）：AI 錯誤地猜測某物是食物（“ 這是意大利辣香腸披薩 ”），儘管實際上這是一張你二表弟的臉的照片（曾經在你母親那邊被移除），因為他在上次訪問夏威夷群島時忘記塗防曬霜。

False Negative (FN): AI incorrectly guessed that something is food (“ This is a pepperoni pizza. ”), although in fact it was a picture of your second cousin ’ s face (once removed on your mother ’ s side) after forgetting to put on his sunscreen during his last visit to the Hawaiian islands.)

現在我們的混淆矩陣已經建立，讓我們透過為每個結果分配一個近似的美元值來將其轉換為價值矩陣。

Now that our Confusion Matrix is created, let ’ s convert it into a Value Matrix by assigning an approximate dollar value to each outcome.

真陽性 （TP）：正確猜測食物，可以節省大約 1 分鐘在手機上手動輸入數據的時間，其中包括查找食物然後輸入食物。鑑於此時美國的年薪中位數約為 60,000 美元，約為 30 美元/小時，我們得到的費率約為每分鐘 50 美分。因此，對於每個 TP，用戶節省大約 1 分鐘的工作時間，或相當於 50 美分。

True Positive (TP): Guessing the food correctly, saves about 1 minute of time doing manual data entry on your phone, which includes finding the food and then entering it. Given that at this time, the median yearly wage in the United States is around $60,000, which is about $30/hour, we get a rate of about 50 cents a minute. So for each TP, the user saves approximately 1 minute of work, or the equivalent of 50 cents.

真陰性 （TN）：正確猜測某物不是食物值得一笑，但僅此而已。大多數人不會有意識地浪費時間試圖將他們的親戚作為食物組。（除非你是漢尼拔 · 萊克特。因此，我們同樣可以對 TN 進行估值，約為 50 美分。

True Negative (TN): Guessing correctly that something is not food is worth a chuckle, but not much more than that. Most people will not consciously waste time trying to enter their relatives as food groups. (Unless you are Hannibal Lecter maybe.) So we can value TN likewise at around 50 cents.

誤報 （FP）：也稱為猜錯食物，很煩人，因為顧客花了精力拍照並上傳，現在必須手動輸入食物，所以這種煩惱可能會花費我們從一開始手動輸入食物所花費的時間的兩倍。因此，讓我們稍微任意地將此結果分配為 －1 美元的值（或 1 美元的費用）。

False Positive (FP): Also known as guessing the food incorrectly, is annoying, because the customer invested effort into taking a picture and uploading it, as well as now having to enter the food manually, so the annoyance perhaps costs us double the amount of time spent entering the food manually from the start. So let ’ s somewhat arbitrarily assign this outcome the value of －$1 (or expense of $1).

假陰性 （FN）：也稱為錯誤地猜測表弟的臉是食物，這也不是正常人會浪費時間的事情，但一旦發生，它會削弱信任，因此賦予它一些價值很重要。我猜這個結果比將燕麥片誤認為小麥奶油嚴重 5-10 倍。因此，讓我們為其分配 －10 美元的值（或 10 美元的費用）。

False Negative (FN): Also known as incorrectly guessing your cousin ’ s face is food, is again not something normal people will be wasting their time with, but it would erode trust when it occurs, so it ’ s important to assign some value to it. I would guess this outcome is about 5 – 10 times more egregious than mistaking oatmeal for Cream of Wheat. So let ’ s assign it the value of －$10 (or expense of $10).

現在你已經明白了：你必須做出至少兩個正確的真陽性猜測（每個 0.50 美元）來抵消每個錯誤的誤報猜測（-1 美元）。

So there you have it: you have to make at least two correct true positive guesses (at $0.50 each) to counteract each incorrect false positive guess (at －$1).

回想一下，我們的準確性公式是：

Recall that our formula for accuracy is:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

鑑於在正常使用期間我們基本上可以忽略真陰性和假陰性計數，因此在我們的例子中簡化了準確性 = （TP） / （TP + FP）。假設 TP = 2 FP，

Given that we can mostly ignore true negative and false negative counts during normal use, simplified accuracy in our case = (TP) / (TP + FP). Given that TP = 2 FP,

Accuracy = (TP) / (TP + 2TP) = TP/3TP

或者在總共猜測 100 次時，100 / 3 * 100 = 100 / 300 = 0.33 = 33%。這意味著我們的人工智慧的準確率必須大於 33%。這其實是相當低的！這意味著對於這個用例，我們的人工智慧在猜測食物類型方面可以非常積極。

Or at 100 total guesses, 100 / 3 * 100 = 100 / 300 = 0.33 = 33%. Which means that our AI has to have an accuracy greater than 33 percent. This is actually quite low! This means that for this use case, our AI can be quite aggressive in guessing the food type.

用棒球來比喻，一個準確的人工智慧會非常努力地不進行好球，因此除非它知道自己可以乾淨利落地擊球，否則它往往不會擊球。相比之下，一個積極進取的人工智能會努力不錯過任何潛在的得分機會，因此它往往會在每個投球中擊球。

To use the baseball analogy, an accurate AI will try very hard not to have a strike, so it tends to not hit unless it knows it can make a clean hit. In contrast, an aggressive AI will try hard not to miss any potential opportunities to score, so it tends to hit at every pitch.

還有一點值得注意：在這種情況下，誤報是漸進性的。這意味著前幾次人工智慧猜錯了食物類型，使用者可能會讓它發出輕微的抱怨。然而，如果人工智慧一直將顧客早上的黑咖啡稱為巧克力棒，用戶的耐心很快就會耗盡，他們很可能會對整個產品認輸。因此，雖然初始準確度不是特別重要（可能低至 33%，例如，多達每兩個猜測中就有一個可能是錯誤的），但快速提高經常輸入的食物的準確度至關重要，這些食物是特定用戶飲食的主食。保持 33% 準確率且不會隨著使用而改進的應用程式可能會在 3-5 次使用內失敗。在使用者結束應用程式之前，有多少誤報是可以的？各種結果的價值是什麼？如果不為每位客戶創建客製化的人工智慧模型，就能實現這種程度的改進嗎？回答這些問題和許多其他相關問題正是使用者體驗如何為人工智慧驅動的專案增加巨大價值。

There is one more point of note: A false positive in this case is progressive. That means the first few times the AI guesses the food type incorrectly, the user will likely let it slide with a minor grumble. However, if the AI will keep calling the customer ’ s morning black coffee a chocolate bar, the user ’ s patience will very quickly wear out and they will likely throw in the towel on the entire product. For this reason, while the initial accuracy is not particularly important (and can be as low as 33 percent; e.g., as many as one out of every two guesses can be wrong) it ’ s going to be critical to rapidly increase the accuracy for frequently entered foods that are a staple of the particular user ’ s diet. The app that maintains 33 percent accuracy and does not improve with use will likely fail within 3 – 5 uses. How many false positives are okay before the user quits the app? What is the value of various outcomes? Can that level of improvement be achieved without creating a bespoke AI model for every customer? Answering these and many other related questions is exactly how UX adds tremendous value to the AI-driven project.

現在輪到你了。針對您自己的使用案例執行此價值矩陣分析練習。（甚至不要想跳過這個練習，並記住將所有液體放在一個加侖大小的袋子裡——稍後會進行二次檢查！

Now it ’ s your turn. Perform this Value Matrix analysis exercise for your own use case. (Don ’ t even think of skipping this exercise, and remember to keep all your liquids in a gallon-sized bag—there will be a secondary inspection later!)

參考

References

1. 1. Kaggle 數據科學競賽以創建準確的 AI 模型（nd）。www.kaggle.com/competitions1.　　Kaggle data science competition to create accurate AI models (n.d.). www.kaggle.com/competitions

2. 2. Gartner 數據與分析高峰會展示，2019 年 3 月 21 日。https://youtu.be/XA2FhDo3hm42.　　Gartner Data and Analytics Summit Showcase, March 21, 2019. https://youtu.be/XA2FhDo3hm4

3. 3. 森古普塔，A.（2019）。過度炒作人工智慧的危及生命的後果。連線。www.wired.com/story/the-life-threatening-consequences-of-overhyping-ai3.　　Sengupta, A. (2019). The life-threatening consequences of overhyping AI. Wired. www.wired.com/story/the-life-threatening-consequences-of-overhyping-ai

4. 4. 納拉揚達斯，D. 和森古普塔，A. （2023）。使用人工智慧在動盪的世界中調整您的行銷和銷售。哈佛商業評論。https://hbr.org/2023/04/using-ai-to-adjust-your-marketing-a

nd-sales-in-a-volatile-world4.　　Narayandas, D., & Sengupta, A. (2023). Using AI to adjust your marketing and sales in a volatile world. Harvard Business Review. https://hbr.org/2023/04/using-ai-to-adjust-your-marketing-and-sales-in-a-volatile-world

5. 5. Sengupta， A. 等人（2023 年）。接受人工智慧挑戰。Aible.com。https://aible.com5.　　Sengupta, A., et al. (2023). Take the AI challenge. Aible.com. https://aible.com

6. 6. 9 月 11 日襲擊（2024 年）的經濟影響。維基百科。https://en.wikipedia.org/wiki/Economic_effects_of_the_September_11_attacks6.　　Economic effects of the September 11 attacks (2024). Wikipedia. https://en.wikipedia.org/wiki/Economic_effects_of_the_September_11_attacks

精確度、召回率和準確性

PRECISION, RECALL, AND ACCURACY

在本章中，我們花了相當多的時間討論精確度、召回率和準確度，但我們尚未正式定義所有術語。在這個稍微「數學」的側邊欄中，我們的目標就是這樣做。

In this chapter, we ’ ve spent a fair bit of time talking about precision, recall, and accuracy, but we ’ ve not yet officially defined all of our terms. In this slightly more “ mathy ” sidebar, we aim to do just that.

精密

Precision

精確度是正確預測 （真陽性） 觀察值與 AI 模型所做出的正面預測總數的比率。它會告訴您模型預測的陽性中有多少是正確的。當誤報成本較高時，高精度至關重要 （1） ：

Precision is the ratio of correctly predicted (true positive) observations to the total number of positive predictions made by the AI model. It tells you how many of the model ’ s predicted positives were correct. High precision is critical when the cost of false positive is high (1):

例如，想像一個充當垃圾郵件過濾器的 AI 模型（因此「這封電子郵件是垃圾郵件」代表正面預測）。精確度將衡量有多少電子郵件實際上是垃圾郵件（真陽性）與被錯誤歸類為

垃圾郵件（誤報）。在這種情況下，非常精確的人工智慧意味著更少的「好」電子郵件被歸類為垃圾郵件。

For example, imagine an AI model that acts as a spam filter (so a “ this email is spam ” represents a positive prediction). Precision would measure how many emails were actually spam (true positive) vs. wrongly classified as spam (false positive). A very precise AI in this case means fewer “ good ” emails were classified as spam.

召回

Recall

正如我們在第 5 章中討論的，召回率（有時稱為靈敏度或真陽性率）衡量「攻擊性」：例如，人工智慧模型在資料集中找到某物的所有相關正面實例的能力。當錯過任何正面因素代價高昂時，高召回率至關重要 （1）。

As we discussed in Chapter 5, recall (sometimes called sensitivity or true positive rate) measures “ aggressiveness ”: For example, an AI model ’ s ability to find all relevant positive instances of something in the dataset. High recall is critical when missing any positive is costly (1).

使用電子郵件垃圾郵件過濾器示例，召回率告訴我們有多少實際垃圾郵件被正確識別為垃圾郵件，以及有多少通過過濾器（誤報）。如果您的 AI 模型具有很高的召回價值，您將收到很少的垃圾郵件。然而，您的人工智慧也可能將大量「好」電子郵件識別為垃圾郵件，因為它會盡量不錯過任何潛在的積極因素。

Using the email spam filter example, recall tells us how many actual spam emails were correctly identified as spam versus how many slipped through the filter (false negatives). If your AI model has a high recall value, you will get very little spam. However, your AI will likely also identify a large number of “ good ” emails as spam, because it tries not to miss any potential positives.

正確

Accuracy

正如我們在第 5 章中詳細討論的那樣，準確的模型會努力避免出錯，因此它們經常“ 把錢留在桌面上 ”。當做出錯誤預測（正面或負面）的成本很高時，高精度至關重要。

As we discussed at length in Chapter 5, accurate models try hard not to be wrong, so they often “ leave money on the table. ” High accuracy is crucial when the cost of making an incorrect prediction (positive or negative) is high.

Accuracy = (true positives + true negatives)/(true positives + true negatives + false positives + false negatives)

或者說得簡單一點,

Or to put it simpler,

Accuracy = correct predictions/total predictions

使用電子郵件垃圾郵件過濾器示例，高精度的模型在將事物識別為垃圾郵件時會很謹慎，並且會抵制這樣做，直到它真正確定電子郵件是垃圾郵件。因此，高度準確的人工智慧可能會讓大量垃圾郵件通過，以確保沒有一封「好」電子郵件被過濾器捕獲。

Using the email spam filter example, a model high in accuracy would be cautious in identifying things as spam and will resist in doing so until it ’ s really sure that an email is spam. As a result, a highly accurate AI may let a lot of spam emails through to ensure none of the “ good ” emails get caught in the filter.

資料科學計量經常相互對抗

Data Science Metrics Often Work Against Each Other

在現實世界中，精確度、召回率和準確性經常相互矛盾。改進一個會犧牲其他一個，因為我們專注於模型效能的不同方面：

In the real world, precision, recall, and accuracy often work against each other. Improving one sacrifices the others, because we focus on a different aspect of model performance:

精確度旨在最大限度地減少誤報：這是當誤報成本較高時您優先考慮的指標。例如，錯誤地指控某人欺詐可能會將無辜者送進監獄。

Precision aims to minimize false positives: It ’ s the metric you prioritize when the cost of a false positive is high. For example, wrongly accusing someone of fraud can send an innocent person to jail.

召回旨在最大限度地減少假陰性：當錯過真陽性可能導致嚴重後果時，這一點至關重要。例如，高召回率有利於診斷可能危及生命但易於安全活檢的癌症。

Recall aims to minimize false negatives: It ’ s essential when missing a true positive can lead to severe consequences. For instance, high recall is good for diagnosing cancers that might be life-threatening but are easy and safe to biopsy.

準確性旨在最大限度地減少所有錯誤：當任何錯誤（誤報或誤報）可能導致嚴重後果時，這一點至關重要。再次使用醫學範例，當它用於識別何時需要切除膽囊的手術時，您需要一個準確的模型，因為您希望絕對確定自己沒有進行不必要的手術（假陽性），同時也不會錯過潛在的危及生命的感染（假陰性）。準確的模型會盡量減少這兩個錯誤。

Accuracy aims to minimize all mistakes: It ’ s essential when any mistake (false positive or false negative) might lead to severe consequences. To use the medical example again, you want an accurate model when it ’ s used to identify when a surgery to remove the gallbladder is needed, because you want to be absolutely sure you are not doing unnecessary surgery (false positive) while also not missing a potential life-threatening infection (false negative). An accurate model tries to minimize both mistakes.

參考

Reference

1. 1. 範奧滕，N.（2024）。機器學習的精確度和召回率變得簡單：如何處理權衡。現貨情報。https://spotintelligence.com/2024/09/11/precision-and-recall1. Van Otten, N. (2024). Precision and recall in machine learning made simple: How to handle the trade-off. Spot Intelligence. https://spotintelligence.com/2024/09/11/precision-and-recall

可變人工智慧模型——為什麼如果猜測價格，準確性仍然是胡說八道

VARIABLE AI MODELS—WHY ACCURACY IS STILL BULLSHIT IF IT ’ S GUESSING THE PRICE

並非所有人工智慧成本和收益都巧妙地融入到特定的 TP/TN 或 FP/FN 結果中。一些現實世界的用例包括人工智慧試圖猜測連續變數（例如價格或時間）的最佳值。在變數最佳化

模型中，資料科學家同樣談論準確性（保守）和召回率（激進）人工智慧模型。讓我們看一個現實生活中的用例，這將有助於證明為什麼人工智慧準確性在可變的人工智慧用例中是一個如此具有破壞性的神話。

Not all of the AI costs and benefits neatly fold into a specific TP/TN or FP/FN outcome. Some real-world use cases include AI trying to guess the optimal value of a continuous variable, such as price or time. In the variable optimization models, data scientists likewise talk about accuracy (conservative) and recall (aggressive) AI models. Let ’ s look at a real-life use case that will help demonstrate why AI accuracy is such a damaging myth in variable AI use cases.

想像一下，您創建了一個人工智慧來幫助人們為待售房屋定價。如果人工智慧收費太低，你就會把錢留在桌面上。如果AI收費過高，你的房子可能需要更長的時間才能賣掉，而且你會在房子空置的情況下賠錢。

Imagine you create an AI that helps people price their house for sale. If AI charges too little, you are leaving money on the table. If AI charges too much, your house may take longer to sell, and you will lose money while the house is standing empty.

在每種情況下你會賺多少錢或損失多少錢？讓我們快速計算一下。

How much money will you make or lose in each case? Let ’ s do a quick calculation.

好吧，假設這棟房子價值一百萬美元。

Well, let ’ s say the house is worth $1 million.

如果您的模型基於召回率並且具有侵略性，那麼它很可能會低估房屋的價格。好消息是房子會很快賣掉，所以你只會損失你低估的金額。假設一個激進的人工智慧將你的房子定價低估了 100,000 美元，所以現在你損失了 100,000 美元，賺了 900,000 美元。

If your model is based on recall and is aggressive, it is likely to underprice the house. The good news is that the house will sell quickly, so you will only lose the amount that you underpriced by. Let ’ s say an aggressive AI underpriced your house by $100,000, so now you lost $100,000 and made $900,000.

另一方面，如果人工智慧非常準確，那麼它很可能把房子的價格高了。所以房子需要更長的時間才能賣掉——比方說，多一年。讓我們來看看本案中應計的機會成本。首先，通過使用準確模型並對房屋定價過高，您將失去 6% 的利息，您可以在 1,000,000 美元（即 60,000 美元）中賺取 1,000 美元。其次，當房子空置時，你本可以賺取租金......因此，12 個月每月 3,000 美元，即 36,000 美元。因此，您總共損失了 60,000 美元 + 36,000 美元或總計 96,000

美元。然而，當價格過高的房子最終售出時——假設它多賣了 100,000 美元——所以你賺了 1,100,000 美元。

On the other hand, if AI was very Accurate, chances are it overpriced the house. So the house would take longer to sell—let ’ s say, one year longer. Let ’ s take a look at the opportunity costs accrued in this case. First off, by using an Accurate model and overpricing the house, you ’ d be out 6 percent interest you could have been earning on the $1 million, which is $60,000. Second, there is rent you could have been earning while the house is standing empty … So that ’ s $3,000/month for 12 months, or $36,000. So you are out a total of $60,000 + $36,000 or $96,000 total. However, when the overpriced house finally sells—let ’ s say it sells for $100,000 more—so you earn $1,100,000.

在我們的最終計算中，1,100,000 美元 － 96,000 美元 = 4,000 美元，因此，對於 Accurate AI，您現在領先 4,000 美元的 Accurate AI 定價過高的房屋的總利潤。

In our final calculation, $1,100,000 － $96,000 = $4,000, so seemingly, with an Accurate AI, you are now ahead $4,000 of total profit on a house that the Accurate AI has overpriced.

不幸的是，在現實世界中，事情很少如此簡單。也許在房子空置一年後，它會增加維護和保險，或者房地產市場下跌，或者你因為房子月復一月地賣不出去而感到緊張。

Unfortunately, things are rarely so simple in the real world. Maybe after one year of the house standing empty and not selling, it racks up maintenance and insurance, or the housing market tanks, or you get sick from the sheer nerve-wracking tension as the house does not sell month after month after month.

哎喲。這些壓力都不值區區 4,000 美元！

Ouch. None of that stress is worth a measly $4,000!

所以這就是人工智慧準確性是胡說八道的原因。以最大程度的召回率為目標的激進人工智慧會降低房屋的價格，因為它將專注於利用每一個銷售機會。事實上，激進的人工智能將通過將房屋定價低到每個潛在買家都想出價來實現最大的召回率。

So here ’ s exactly why AI Accuracy is bullshit. Aggressive AI with a goal of maximum recall would underprice the house as it will focus on taking advantage of every sales opportunity. In fact, aggressive AI will achieve maximum recall by pricing the house so low that every single potential buyer will want to bid on it.

另一方面，Accurate AI 會非常努力地不出錯，因此它的價格可能會過高，因此房子需要更長的時間才能出售。事實上，非常準確的人工智慧會為每一美元而拼盡全力，因此它很可能會將其定價得如此之高，以至於沒有人會考慮競標它。

On the other hand, Accurate AI would work very hard not to be wrong, so it ’ s likely to overprice, so the house takes longer to sell. In fact, very accurate AI will fight hardest for every dollar, so it ’ s likely to price it so high that no one will even think of bidding on it.

現在，在現實世界中，你往往不想要非常「準確」的AI，因為你不想等一年才賣掉你的房子！您也不想要一個非常「激進」且召回率高的人工智慧，因為您不想因定價過低而一開始就損失一大筆錢。相反，您需要一個平衡的人工智慧，它將根據現實世界的考慮因素而不是數據科學指標來優化您的價格，並將您的房子定價在中間標記：1,000,000 美元。這與準確性和回憶力關係不大，而與使用者體驗人員需要訓練自己詢問有關人工智慧預測的問題有關。

Now, in the real world, you often don ’ t want AI that is very “ accurate ” because you don ’ t want to wait a year to sell your house! Nor do you want a very “ aggressive ” AI with high recall because you don ’ t want to lose a bunch of money right out of the gate by underpricing. Instead, you want a balanced AI that will optimize your price based on real-world considerations, not data science metrics and price your house right at the middle mark: $1,000,000. That has very little to do with accuracy and recall and everything to do with the questions UX people need to train themselves to ask about AI predictions.

Accuracy has no value in the real world. Thus when you hear the statement “ This AI is accurate, ” it should sound like bullshit to you because it most certainly is.

當您聽到「這個人工智慧是準確的」時，您應該立即詢問人工智慧預測的成本和影響，因為定價過高或定價過低很少會產生同等的影響。不要只相信資料科學家的話，因為他們通常不使用現實世界的指標。提出好的用戶體驗問題。花時間了解 AI 預測的更廣泛影響，並幫助您的團隊和客戶選擇最適合現實世界的 AI 模型。

When you hear “ This AI is accurate, ” you should immediately ask about the cost and impact of AI predictions, because overpricing or underpricing rarely have equal impact. Don ’ t just take the data scientist ’ s words for it because they typically do not use real-world metrics. Ask good UX questions. Take the time to understand the broader impacts of AI predictions, and help your team and your customers choose the AI model that works best in the real world.

另一個需要考慮的現實點是處理能力。如果一個銷售人員每天可以打電話給 10 個潛在客戶，那麼一個建議兩個潛在客戶的準確 AI 將像建議 100 個潛在客戶的激進模型一樣毫無用處;在每種情況下，建議都會太少或太多！如果客戶只有這麼多時間或注意力（幾乎總是如此），這個問題尤其重要。作為用戶體驗專業人士，您應該找出該特定人員或團隊每單位時間內有多少建議或警報是最佳的，並指導數據科學人員相應地創建 AI 模型。請注意，不同人的處理能力可能不同;例如，積極進取的銷售人員可能每天能夠致電 20 甚至 30 個潛在客戶並完成 5 筆交易，而更有條理和風度翩翩的銷售人員可能最好只用 7 或 8 個潛在客戶來完成相同數量的交易。（因此，我們積極進取的銷售人員可能需要更積極的 AI 模型，而有條不紊的銷售人員可能需要更準確的模型來實現相同的投資回報率。人工智慧模型應盡最大努力透過調整資料科學數字以最適合人類的處理能力來適應每個人的理想偏好。

Another real-world point to consider is processing capacity. If a salesperson can call 10 prospects a day, an accurate AI that suggests two prospects will be as useless as the aggressive model that suggests 100 prospects; in each case, it will be too few or too many suggestions! This problem is especially important if the customer has only so much time or attention to spare (which is pretty much always). As a UX professional, you should find out how many suggestions or alerts are optimal per unit of time for this specific person or team and instruct the data science folks to create the AI model accordingly. Note that processing capacity may be different for different humans; for example, an aggressive salesperson might be able to call 20 or even 30 prospects per day and close five deals, whereas a more methodical and personable salesperson might do best with only seven or eight prospects to close the same number of deals. (Thus our aggressive salesperson might need a more aggressive AI model, and the methodical salesperson might need a more accurate model to achieve the same ROI.) The AI model should do its best to accommodate each human’s ideal preference by adjusting the data science numbers to best suit the human processing capacity.