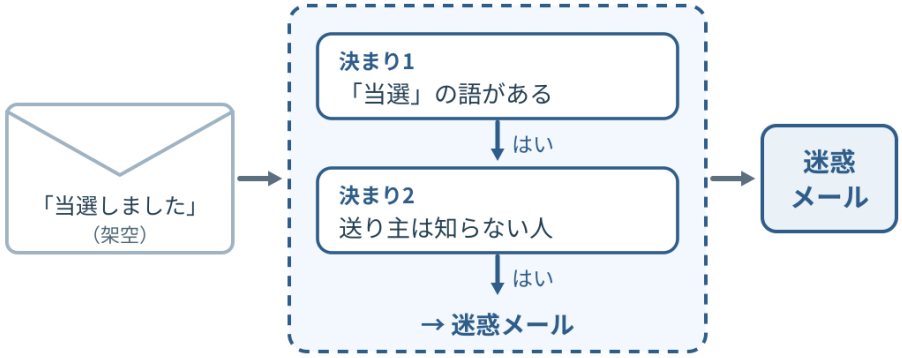
AIの答えは理由が見えにくい（ブラックボックス）

AIは、なぜその答えを出したのかを説明できるのだろうか

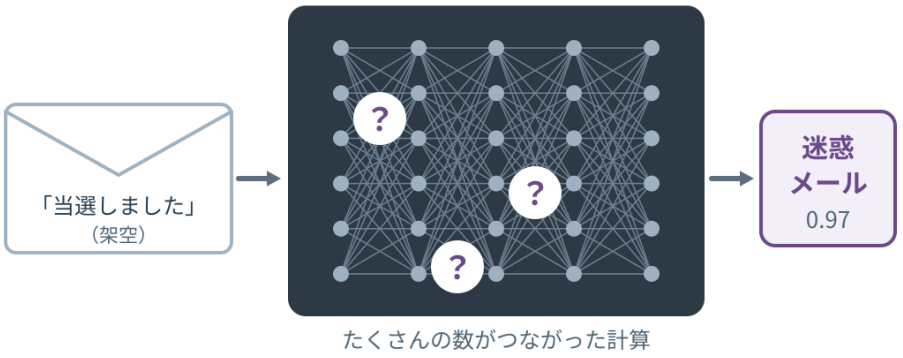
例：迷惑メールを見分ける（架空）

ルールで動くプログラム　中が見える箱



✓ **理由：「当選」の語があり、知らない送り主だから**
人が書いた決まりなので、理由をたどれる

深層学習のAI　中が見えにくい箱



たくさんの数がつながった計算

? **理由：たくさんの数の組み合わせで決まる**
人がたどって説明するのは難しい

決まりを決めるのは	左：人が書く	右：データから学ぶ
よいところと課題	左：理由をたどれる。すべての決まりを書くのは大変	右：決まりを書かなくても見分けられる。理由は見えにくい

説明が求められる場面では　例：進路・お金・健康など、人の生活に大きく関わる判断

- AIの答えだけで決めない
- 人が理由を確かめ、説明できるようにする
- 答えの理由を示す工夫（説明可能性）が研究されている

※ 例は架空。確からしさの数は説明のための例。