Automating the process of screening research articles for living systematic reviews

Ranbir Singh Kochar, MSc Data Science student, Lancaster University. Email: ranbirkochar@gmail.com / r.kochar@lancaster.ac.uk
Rhys Peploe, Statistician, Infectious Diseases Data Observatory, Oxford

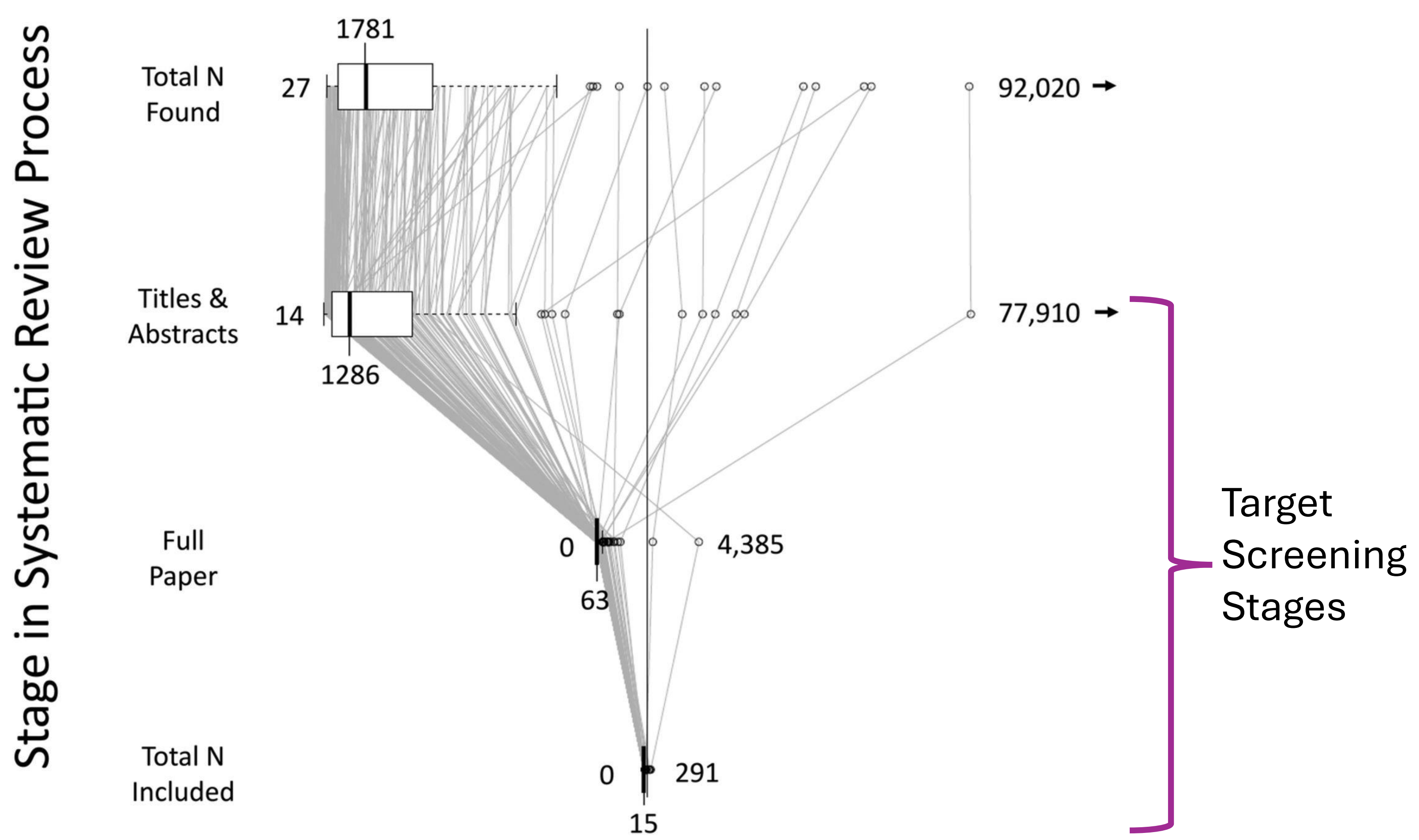
Problem:

Screening articles for systematic review is time consuming!

Median time to conduct and publish a systematic review is 1,110 person hours. It can go up to 2,518 hours. Greatest proportion of this time is used on screening research articles (Borah *et al*., 2017)

This work provides a solution to save this time using machine learning models for the last two screening stages of living systematic review.

Aggregated literature filtration process based on counts reported (n=195).

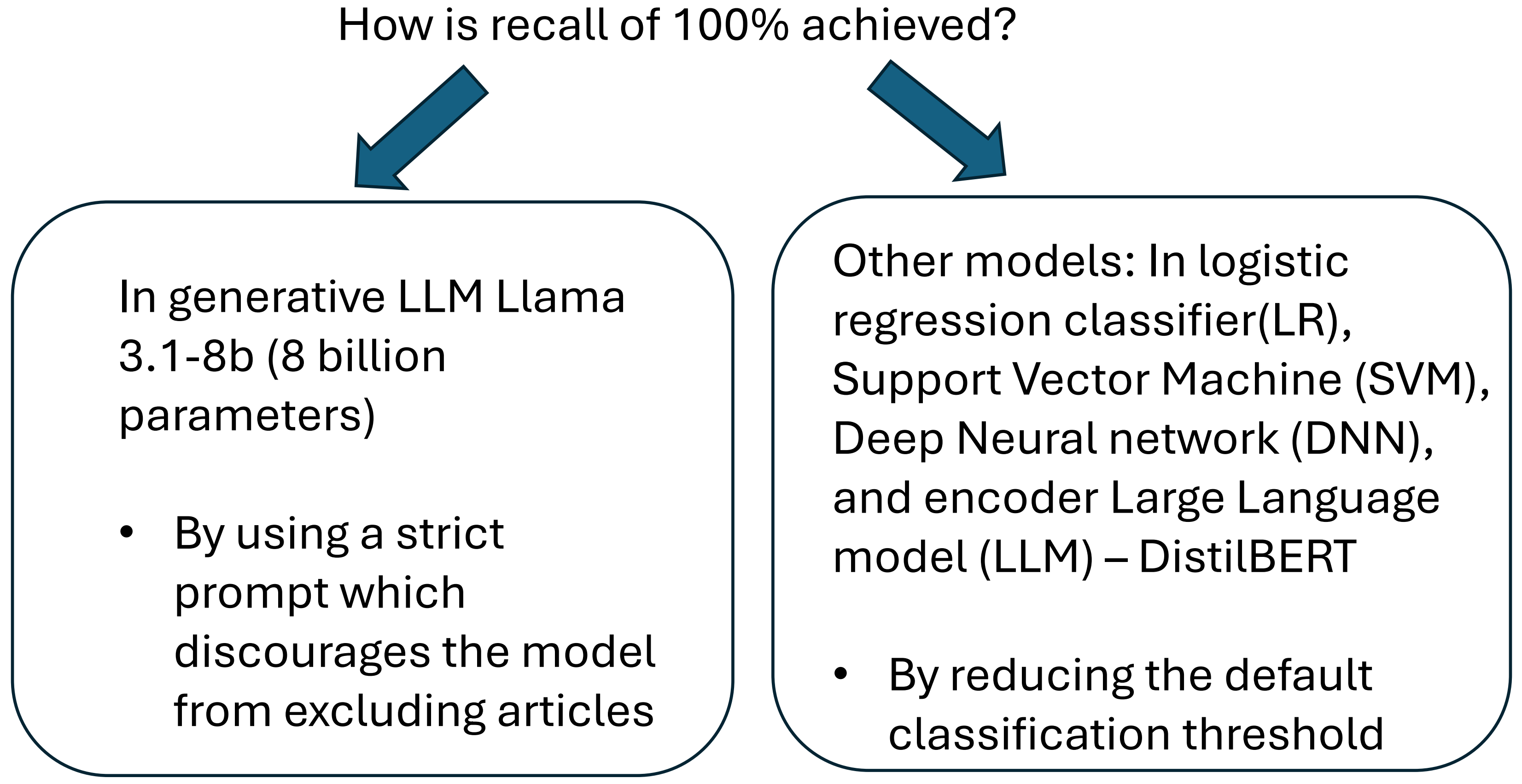


Source: Borah *et al*. (2017)

Method:

Desired recall (sensitivity) = 100% !

If an article that is supposed to be included gets screened out, quality of the systematic review can be severely compromised.



How is recall of 100% achieved?

In generative LLM Llama 3.1-8b (8 billion parameters)

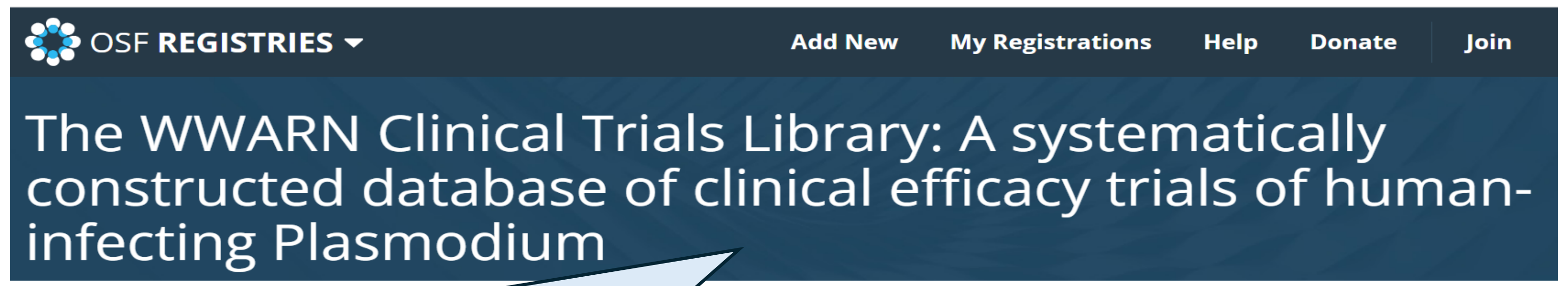
- By using a strict prompt which discourages the model from excluding articles

Other models: In logistic regression classifier(LR), Support Vector Machine (SVM), Deep Neural network (DNN), and encoder Large Language model (LLM) – DistilBERT

- By reducing the default classification threshold

Cost of targeting 100% recall: reduction in accuracy. Still very useful as it saves time without compromising the quality of the review.

Data:



The WWARN Clinical Trials Library: A systematically constructed database of clinical efficacy trials of human-infecting Plasmodium

Downloaded based on database created by Infectious diseases data observatory, Oxford:

11,742 titles+abstracts 1,846 full articles

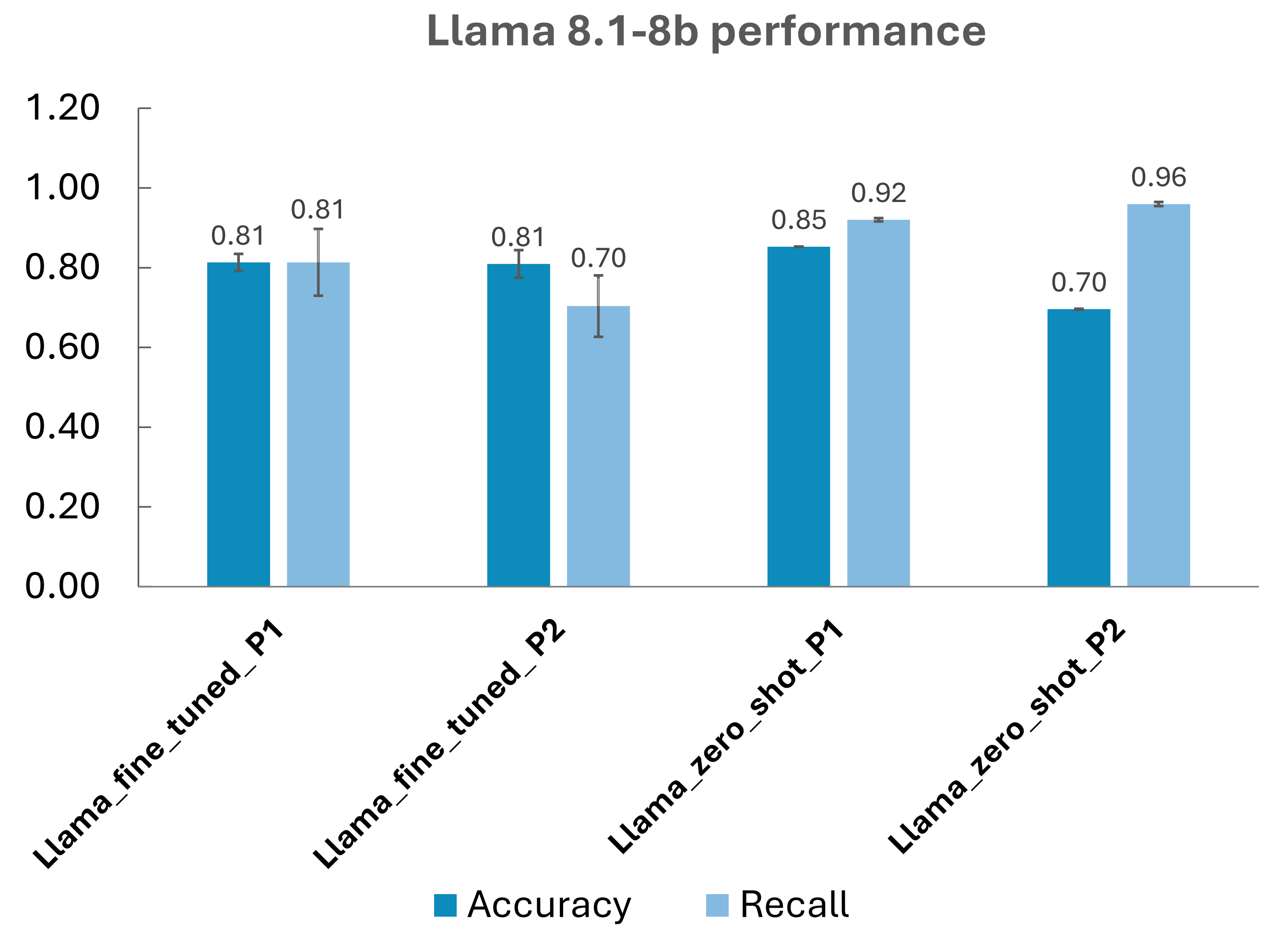
- All models except LLM Llama 3.1 8b were separately trained on each of the two datasets
- Llama 3.1-8b required prompting

Example of a prompt used for fine tuning, and zero-shot classification:

"A systematic review of studies on malaria treatment needs to be conducted. Research articles presenting primary clinical trials on the efficacy of antimalarial drugs must be included. Articles that are literature reviews, case reports, editorials, commentaries, letters, opinion pieces, corrections, or that do not present a primary study on antimalarial drug efficacy should be excluded. Do not easily exclude an article. Exclude it only if there is strong evidence in support of its exclusion. The abstract of a research article is provided. Read it and make a decision. Answer 'Include' or 'Exclude' based on the given abstract of the paper. Here is the abstract:"

Results:

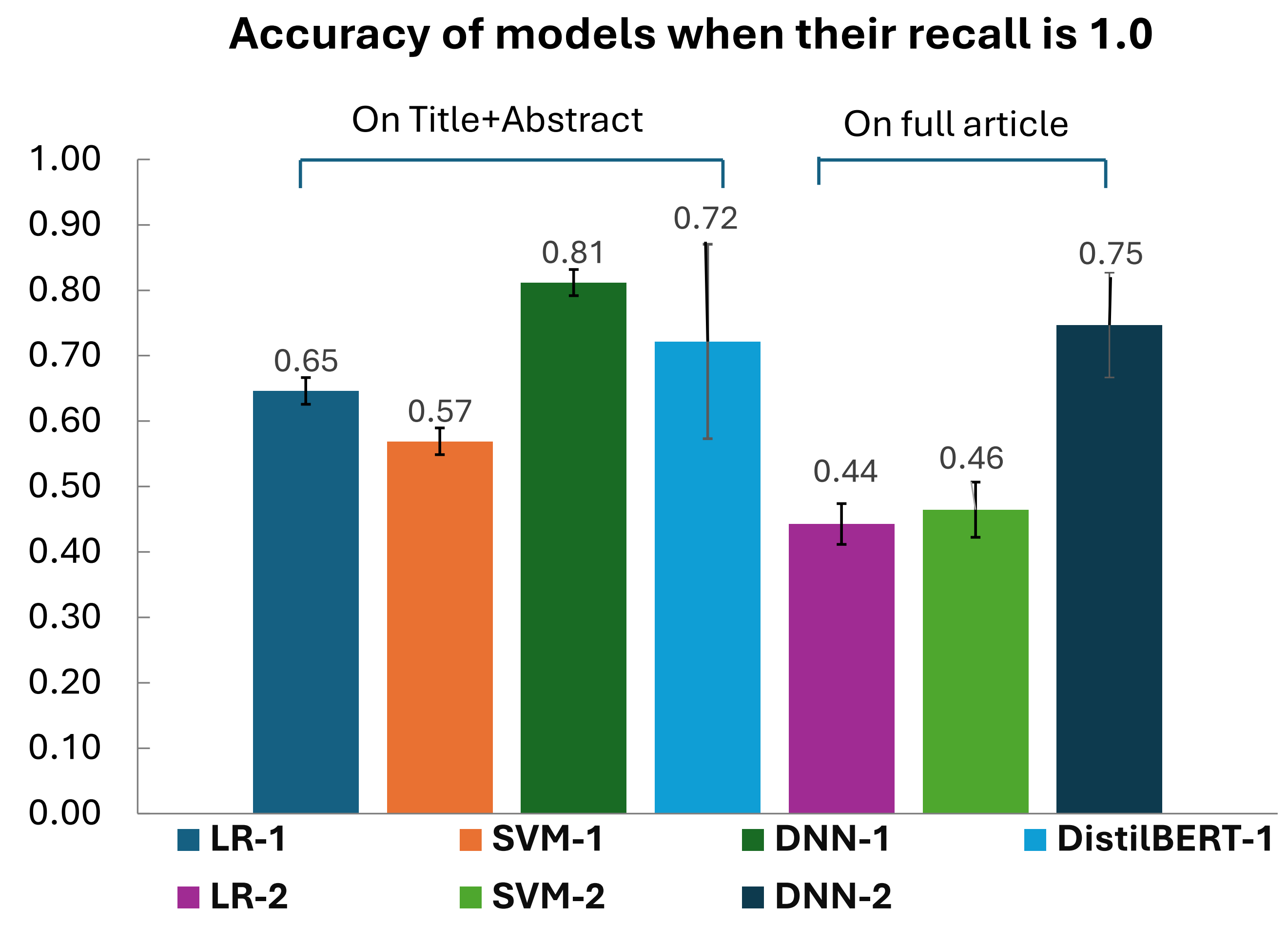
- Llama 3.1-8b



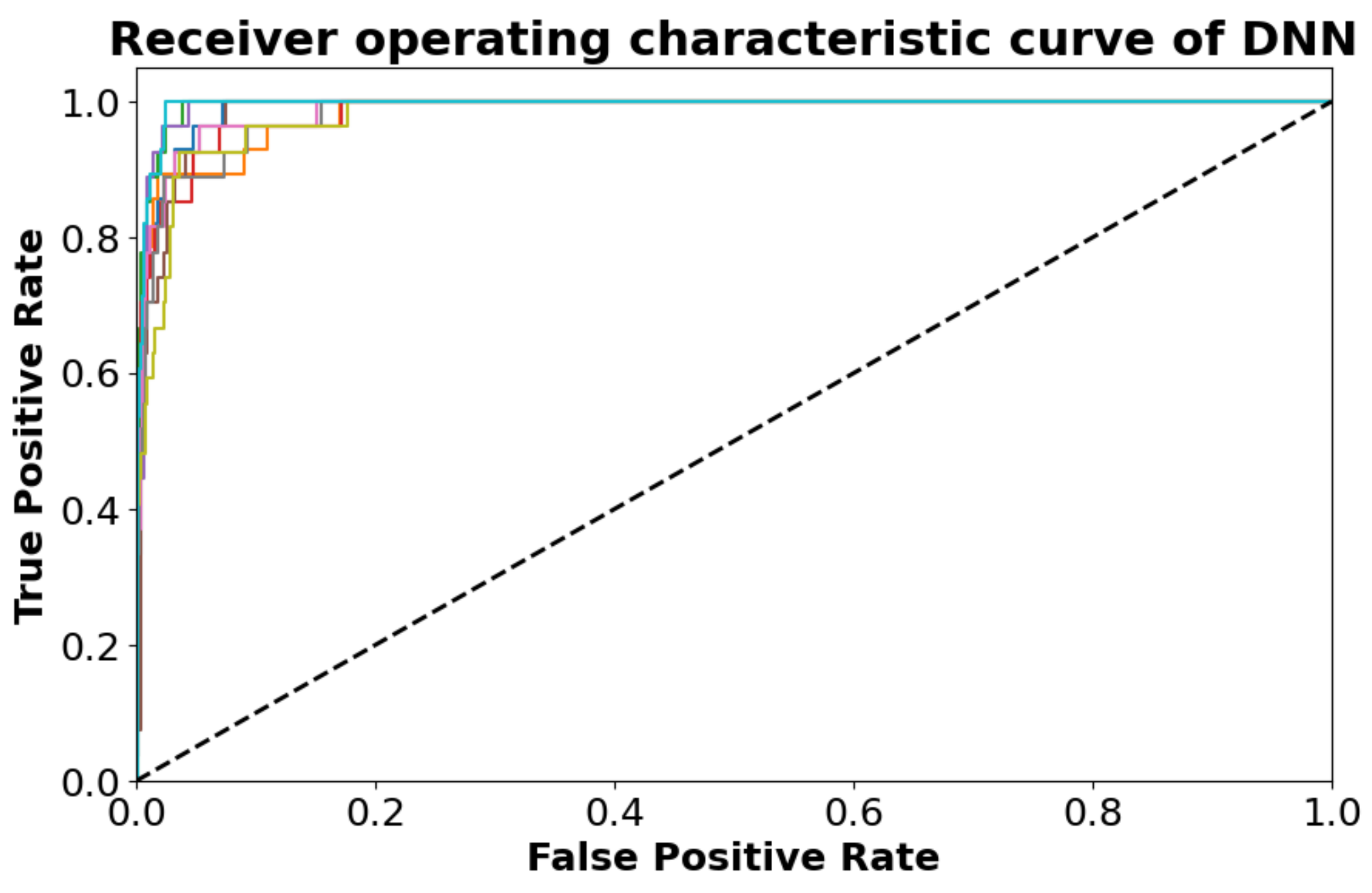
P1 and P2 are two distinct prompts

- Other models

All models were tuned to achieve a recall of 100%. Now it is their accuracy that determines which one is better.



Best model: Deep Neural Network (DNN) trained on Title + Abstract of research articles



DNN performance:

Recall = 1.00 Accuracy = 0.82 Area under ROC curve =0.99

	Predicted Exclude	Predicted Include
Actual Exclude	9,259	2,210
Actual Include	0	273

Implication:

Up to 80% of time saved on the last two stages of screening research articles for a living systematic review update.

Reference: Borah R *et al. (2017)*. *BMJ Open*. 2017;7(2):e012545.