“On the unknowable limits to prediction”

Presented at Oxford LLMs: A Workshop for Social Science Researchers

26th September, 2025

Charles Rahal and Jiani Yan

“We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, … nothing would be uncertain.”

– Laplace (1814), ‘Essai philosophique sur les probabilites’

Imagine you have a prediction task…

It could be predicting the next token, or it could be through asking an LLM to predict some other target feature.

Here is your model performance:

ROC-AUC = 0.85

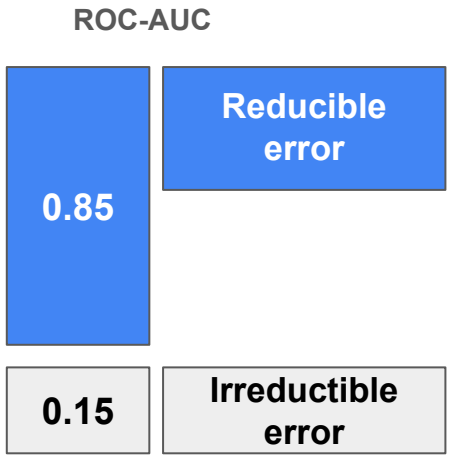
And we know that, the best possible ROC-AUC=1, where we have every prediction perfect

Then…

1. **What's the 0.15, the gap between your current prediction and the best possible prediction?**

2. **Are we ever able to achieve the 'perfect' score as defined by any common metric?**

Existing error decomposition (e.g., Hastie et al. 2001)



From ▭ to ▬ :

What determines accuracy?

- Algorithms
- Variable exclusion/inclusion.
- Construction of the target feature

What is 'irreducible' today may not be irreducible tomorrow!

Everything is EVOLVING!

Including data, algorithms and the construction of the target feature itself

How can we better systematically think about prediction errors?

This is increasingly important as social scientists are increasingly making statements about how ‘predictable’ things are.

This is increasingly important as social scientists are increasingly making statements about how ‘predictable’ things are.

We argue that they should not!

nature computational science

Correspondence | Published: 04 March 2025

On the unknowable limits to prediction

Jiani Yan ✉ & Charles Rahal ✉

Nature Computational Science **5**, 188–190 (2025) | Cite this article

Conceptualise error differently:

Aleatoric error

Inherent randomness
can never be modelled

Epistemic error

In theory eliminable
(resulting from a lack of knowledge)

A typical prediction task workflow

1. **Form a research question:** e.g., *'How predictable is death?'*
2. **Read relevant paper & theories, identify covariates:** *e.g., social determinants of health!*
3. **Find appropriate/ collect data for the research question:** *e.g., biobanks!*
4. **Use proper algorithm(s) to run the prediction task:** *e.g., an LLM!*
5. **Analyse outcomes:** *e.g., any of your favourite model evaluation metrics.*

Once data, research object and covariates are specified, the research design is confirmed; The aleatoric error of the research object target feature becomes fixed.

Table 2: Evaluation of Model Performance in Death Prediction over Ten Seeds

	HRS			SHARE			ELSA		
Metrics	SL	LightGBM	LR	SL	LightGBM	LR	SL	LightGBM	LR
IMV	0.192	0.195	0.111	0.073	0.077	-0.055	0.056	0.060	0.068
ROC-AUC	0.814	0.816	0.823	0.789	0.795	0.453	0.823	0.828	0.837
PR-AUC	0.687	0.691	0.702	0.508	0.522	0.172	0.491	0.499	0.521
Pseudo R^2	0.275	0.280	0.295	0.195	0.207	-0.041	0.208	0.221	0.243
FFC R^2	0.602	0.605	0.613	0.443	0.451	0.279	0.394	0.404	0.421
Test IP	0.317	0.317	0.317	0.202	0.202	0.202	0.155	0.155	0.155

Is 0.823 the predictive ceiling? Can I ever improve it?

Example taken from Yan, J. (2025), '*Revisiting the social determinants of health with explainable AI: a cross-country perspective*', forthcoming at AJE

$$
\begin{aligned}
y_{\text{true}} &= \underbrace{f^*(\mathbf{x}_{\text{true}})}_{\substack{\text{Predictive}\\ \text{ceiling}}} + \underbrace{\varepsilon}_{\substack{\text{Aleatoric}\\ \text{error}}} \\
&= \underbrace{[f^*(\mathbf{x}_{\text{true}}) - f(\mathbf{x}_{\text{true}}|y_{\text{true}})]}_{\substack{\text{Model approximation gain}\\ \text{(Epistemic)}}} \\
&+ \underbrace{[f(\mathbf{x}_{\text{true}}|y_{\text{true}}) - f(\mathbf{x}_{\text{true}}|y_{\text{observed}})]}_{\substack{\text{Measurement gain from } y\\ \text{(Epistemic)}}} \\
&+ \underbrace{[f(\mathbf{x}_{\text{true}}|y_{\text{observed}}) - f(\mathbf{x}_{\text{observed}}|y_{\text{observed}})]}_{\substack{\text{Measurement gain from } \mathbf{x}\\ \text{(Epistemic)}}} \\
&+ \underbrace{y_{\text{predicted}}}_{\substack{\text{Current prediction}}} + \underbrace{\varepsilon}_{\substack{\text{Irreducible}\\ \text{(Aleatoric)}}}.
\end{aligned}
$$

Algorithm

Y — Construct Validity

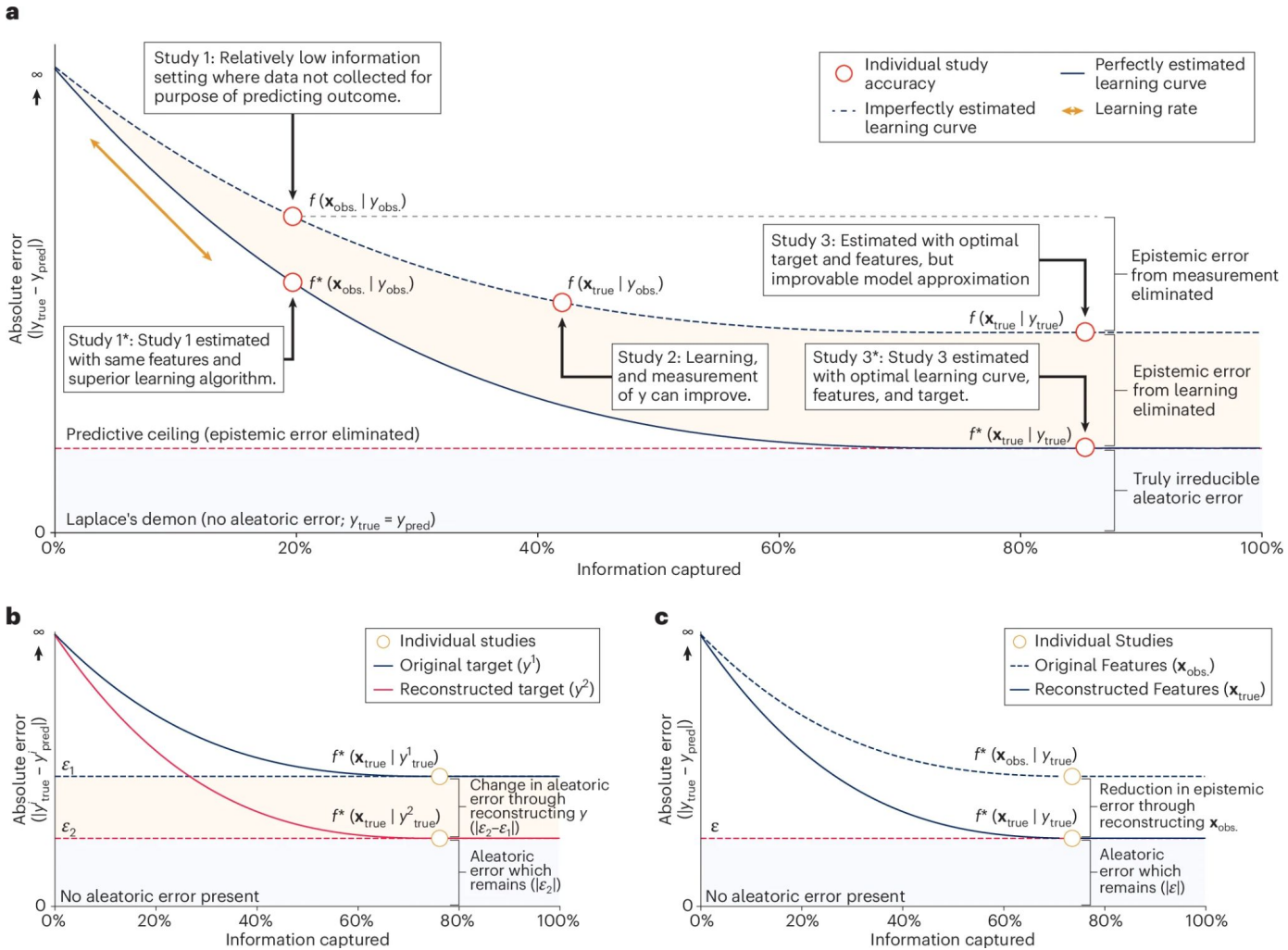
Representation

X — Relevant information inclusion

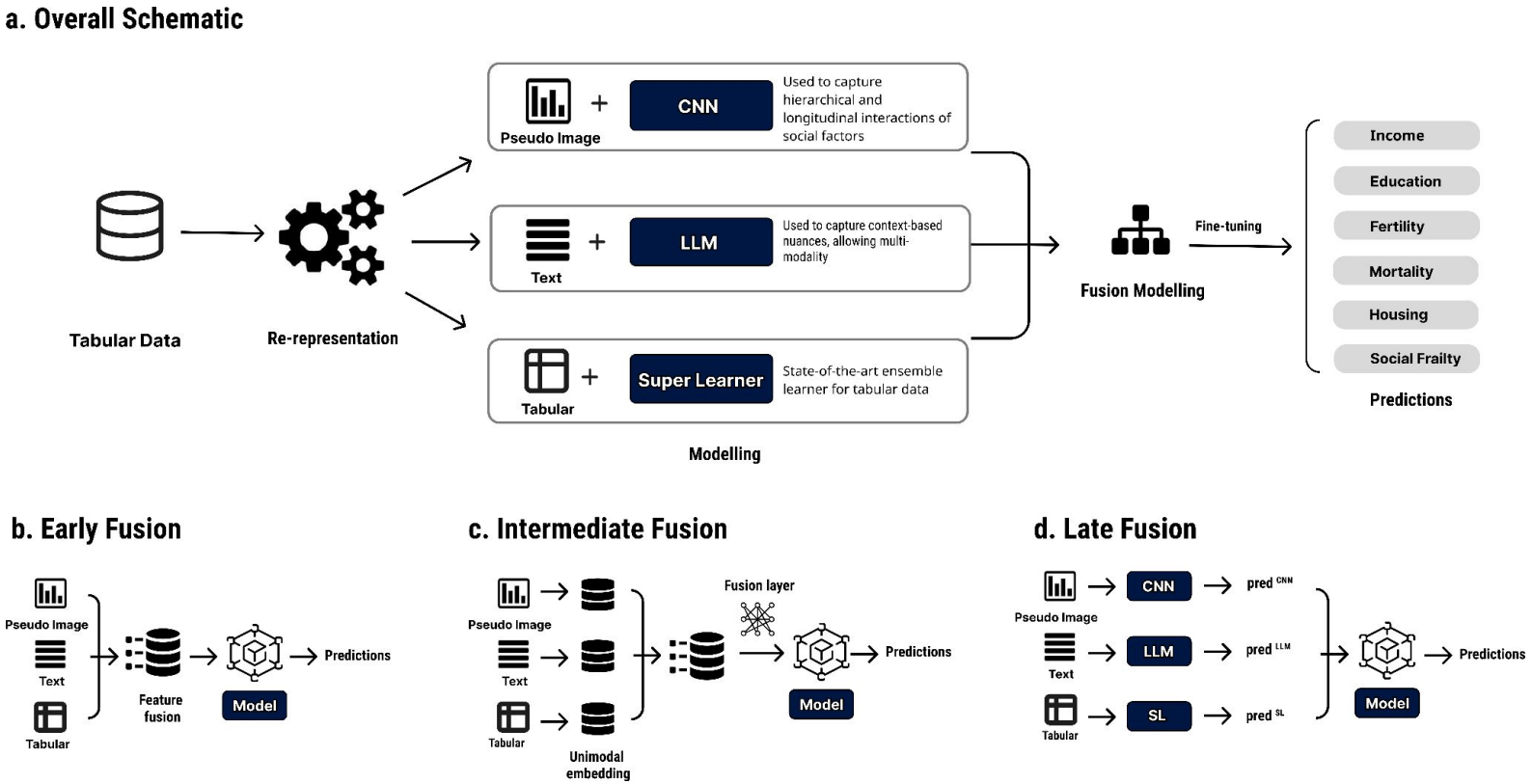
We claim that people should be responsibly using terms while thinking about the limits of human endeavour:

We can only say how predictable things are *conditional* on features and predictive machines.

We emphasize the importance – and develop a new style – of ‘learning curve’.


a
Study 1: Relatively low information setting where data not collected for purpose of predicting outcome.
Individual study accuracy
Perfectly estimated learning curve
Imperfectly estimated learning curve
Learning rate
Absolute error (|y_true − y_pred|)
∞
0
f (x_obs. | y_obs.)
f* (x_obs. | y_obs.)
f (x_true | y_obs.)
Study 3: Estimated with optimal target and features, but improvable model approximation
f (x_true | y_true)
Epistemic error from measurement eliminated
Study 1*: Study 1 estimated with same features and superior learning algorithm.
Study 2: Learning, and measurement of y can improve.
Study 3*: Study 3 estimated with optimal learning curve, features, and target.
Epistemic error from learning eliminated
Predictive ceiling (epistemic error eliminated)
f* (x_true | y_true)
Truly irreducible aleatoric error
Laplace's demon (no aleatoric error; y_true = y_pred)
0%
20%
40%
60%
80%
100%
Information captured
b
Individual studies
Original target (y¹)
Reconstructed target (y²)
Absolute error (|y_true − y_pred|)
∞
0
f* (x_true | y¹_true)
ε1
Change in aleatoric error through reconstructing y (|ε2−ε1|)
f* (x_true | y²_true)
ε2
Aleatoric error which remains (|ε2|)
No aleatoric error present
0%
20%
40%
60%
80%
100%
Information captured
c
Individual Studies
Original Features (x_obs.)
Reconstructed Features (x_true)
Absolute error (|y_true − y_pred|)
∞
0
f* (x_obs. | y_true)
f* (x_true | y_true)
Reduction in epistemic error through reconstructing x_obs.
ε
Aleatoric error which remains (|ε|)
No aleatoric error present
0%
20%
40%
60%
80%
100%
Information captured

What can we do? Comprehensive algorithms and datasets



Can we ever know where we are in the process?

- We have up to seven new approaches for approximating epistemic error across data, model, and measurement.
- These methods include learning curves (Mohr & van Rijn, 2021), Taylor expansions with the law of propagation of uncertainty, and bootstrap variants.
- We also developed a large-scale integrated simulation bench that uses user-defined data-generating mechanisms.

- Aleatoric error is unmodelable; fixed at the point of confirming your research question.
- Aleatoric error levels can vary depending on the research object; some topics are easier to predict than others.
- Epistemic error can be eliminated through
  - Better capture of information
    - Better representativity
    - Better variable construction
    - Better relevant information inclusion
  - Better algorithms

“ Until we have reason to believe that we have eliminated error through better measurement and construction of features and outputs — in addition to the elimination of learning error — we should only tentatively discuss how predictable things are with current data and technology.”