



Sometimes?



Always!

When is it Rational to Review for Privilege?

Douglas W. Oard, University of Maryland (USA)

Jyothi Vinjumur, University of Maryland (USA)

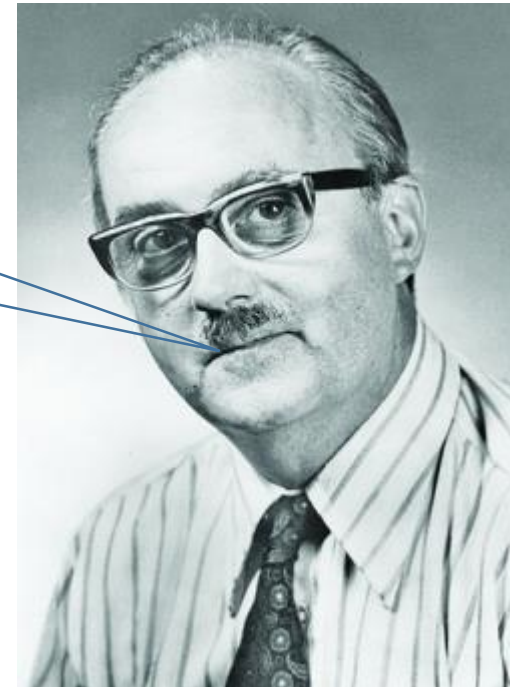
Fabrizio Sebastiani, Italian National Council of Research (Italy)

Rational:

Based on or in accordance with reason or logic

<https://en.oxforddictionaries.com/definition/rational>

All models are wrong
but some are useful



George Box

Some Modeling “Assumptions”

- People don't make mistakes
- Machines do make mistakes
 - Mistakes are more likely on some documents than others
 - Machines know how likely they are to be correct $p(N/P)$ $p(P/N)$
- People have to be paid for their time \$5
- Machine time is free
- Mistakes are costly $c(N/P)$ $c(P/N)$

A “Model”

Expected cost of saying P when it is N = $p(N|P) * c(N|P)$

Expected cost of saying N when it is P = $p(P|N) * c(P|N)$

Review

1: P?  $p(N|P) * c(N|P) = \$50$

2: N?  $p(P|N) * c(P|N) = \$32$

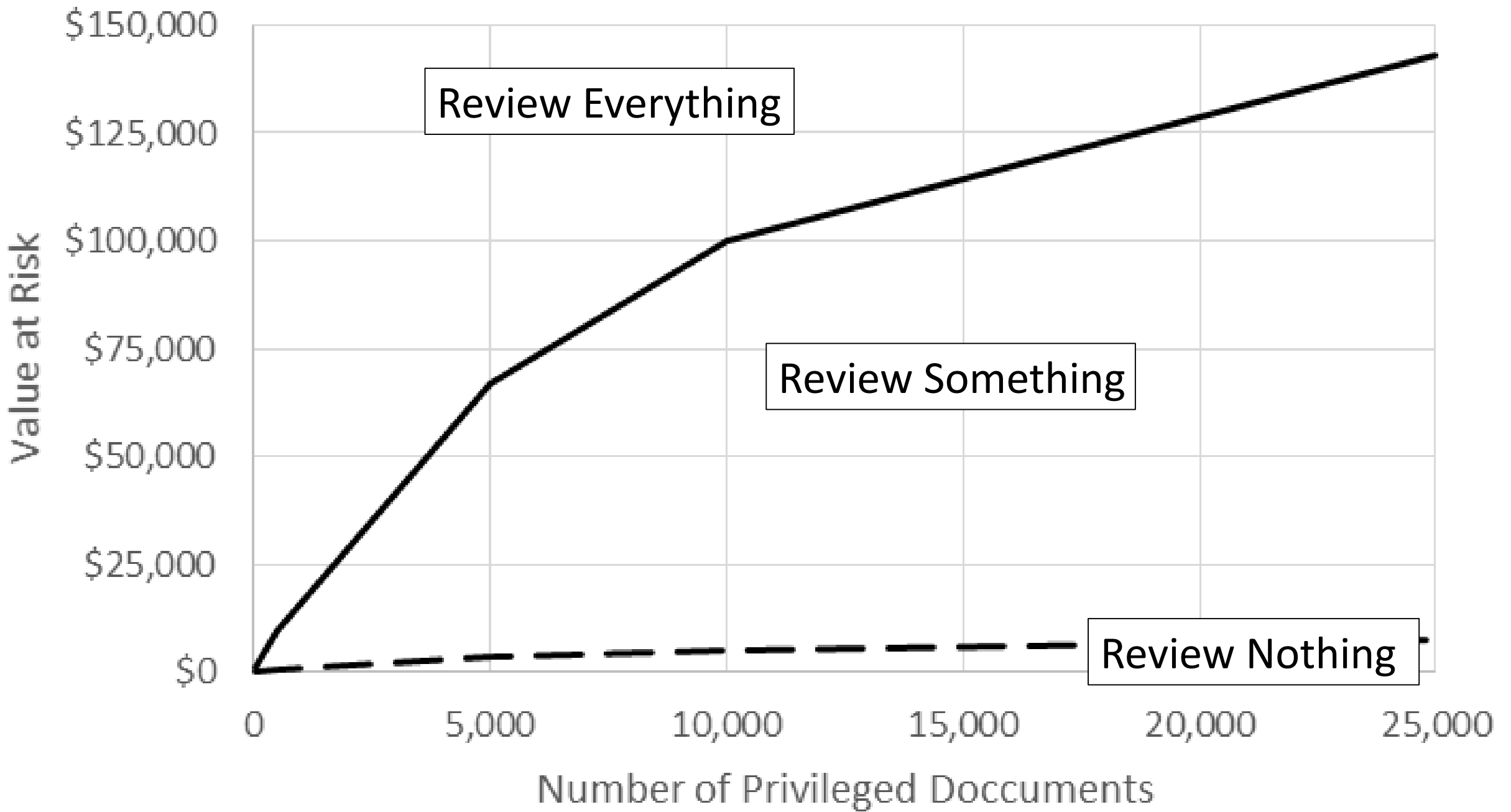
3: P?  $p(N|P) * c(N|P) = \$11$

4: P?  $p(N|P) * c(N|P) = \$6$

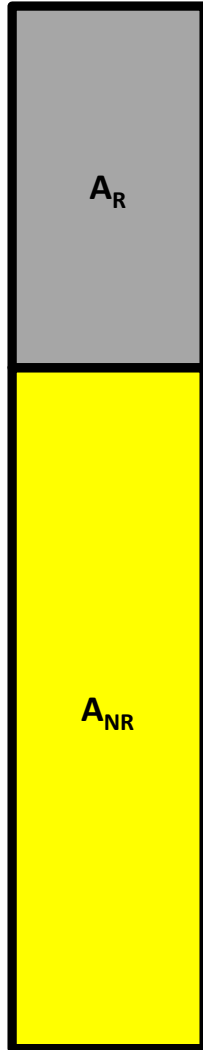
Don't Review

5: N?  $p(P|N) * c(P|N) = \$4$

6: P?  $p(N|P) * c(N|P) = \$1$

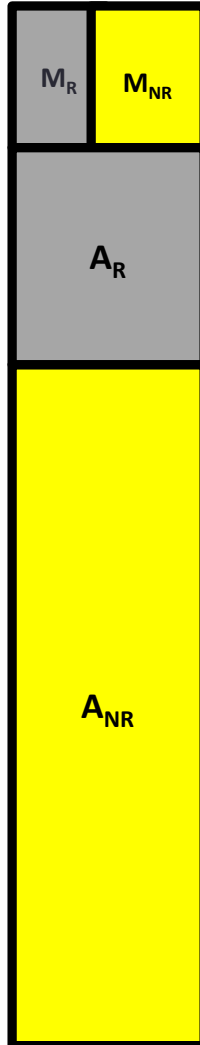


Cost-Sensitive Review



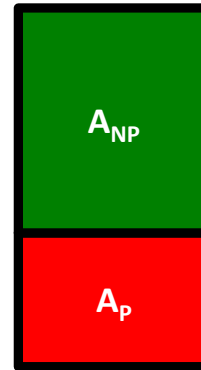
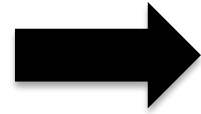
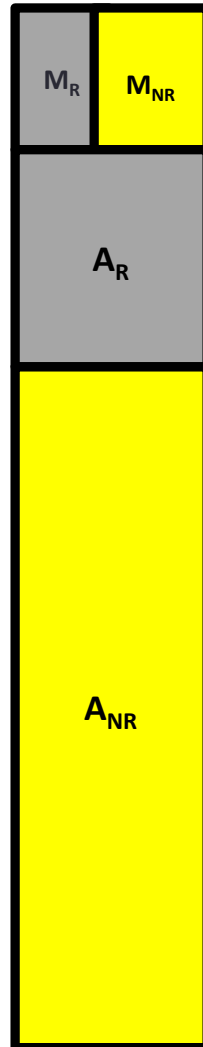
Cost-Sensitive Review

Relevance
Review



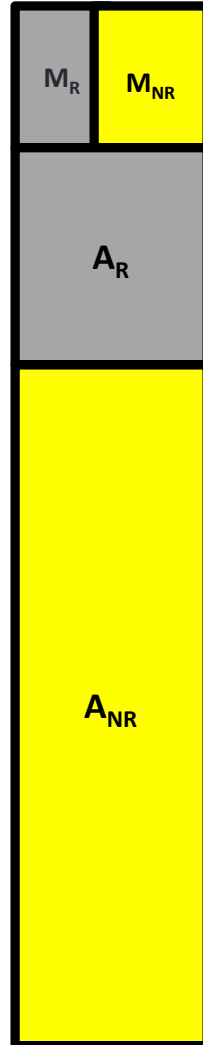
Cost-Sensitive Review

Relevance
Review

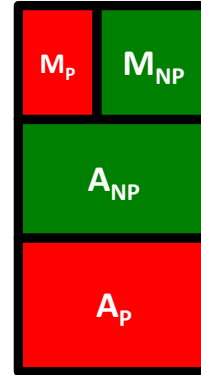


Cost-Sensitive Review

Relevance
Review

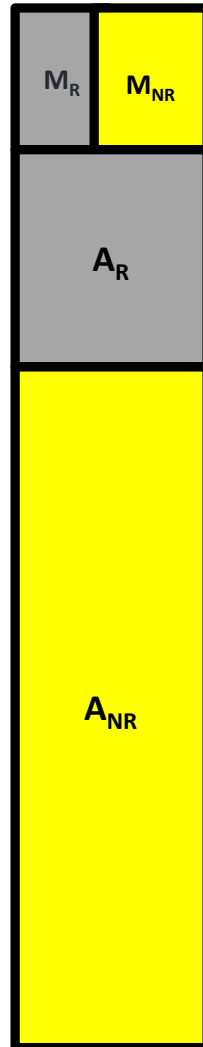


Privilege
Review

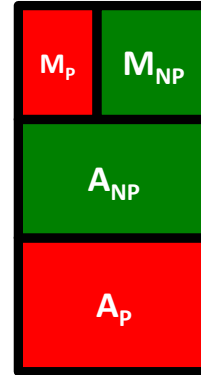


Cost-Sensitive Review

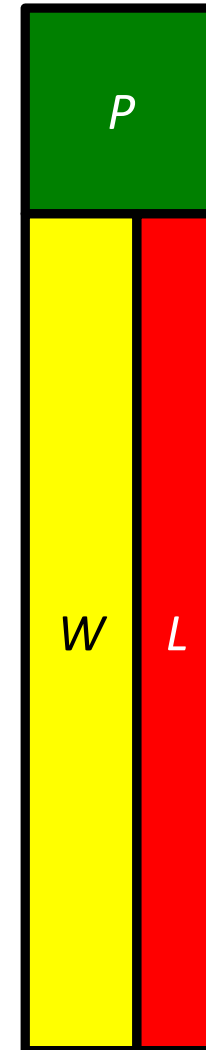
Relevance
Review



Privilege
Review



Production



Risk Minimization Model Inputs

PRODUCE class - Documents that are responsive and not privileged.

LOG class - Documents that are responsive and privileged.

WITHHOLD class - Documents that are not responsive.

Enter your name (optional)

Enter the cost (in USD) of manually reviewing a single document for Responsiveness

1.0

Enter the cost (in USD) of manually reviewing a single document for Privilege

5.0

Please enter the 6 Misclassification Cost Values in the below grid (all values in USD)

		TRUTH		
		PRODUCE	LOG	WITHHOLD
PREDICTIONS	PRODUCE	0	1000	100
	LOG	500	0	200
	WITHHOLD	750	100	0

Submit

Review Condition Results For The Given Cost Matrix

Collection Statistics:

Data: RCV1-v2 Test Collection

Responsive class: Documents in DOMESTIC POLITICS (GPOL) category

Privileged class: Documents in ECONOMICS (ECAT) category

Total number of Documents in the Test Set: 20,000

Responsive class Prevalence: 15.27%

Privileged class Prevalence: 7.44%

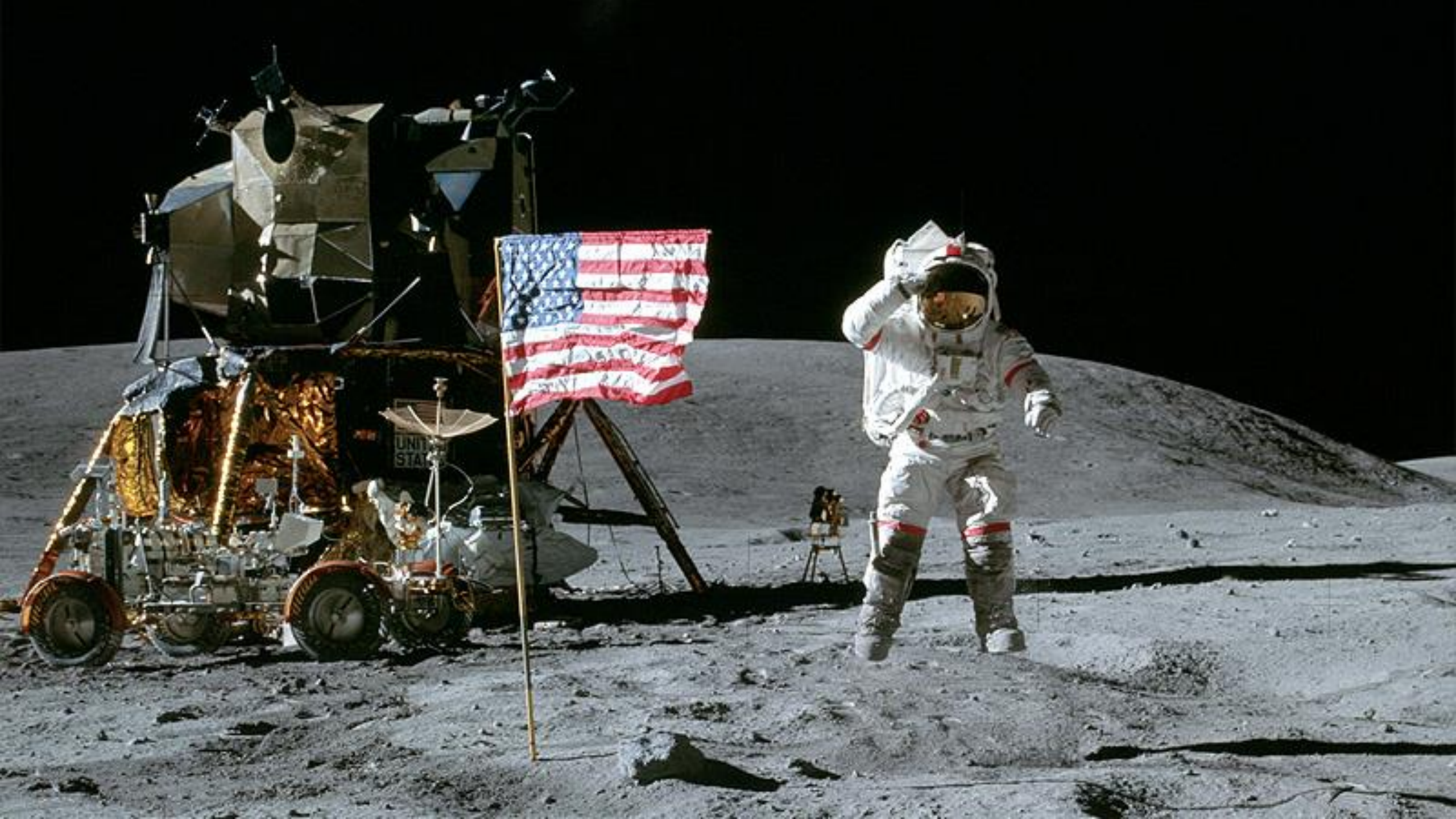
Prevalence of the Intersection among Responsive and Privileged classes: 1.45%

Manual Review Statistics:

Number of Documents to be reviewed for Responsiveness = 17535 of 20,000 (87.675%)

Number of Documents to be reviewed for Privilege = 991 (4.955% of 20,000)





Some Final Thoughts

- Models can provide a rational basis for decision making
 - But they can't make the decision for you
- Sensitivity analysis can explore effect of changing your assumptions
- The model's result can inform your decisions
- The model can also facilitate discussion with other parties