

Minimum Viable Disclosure

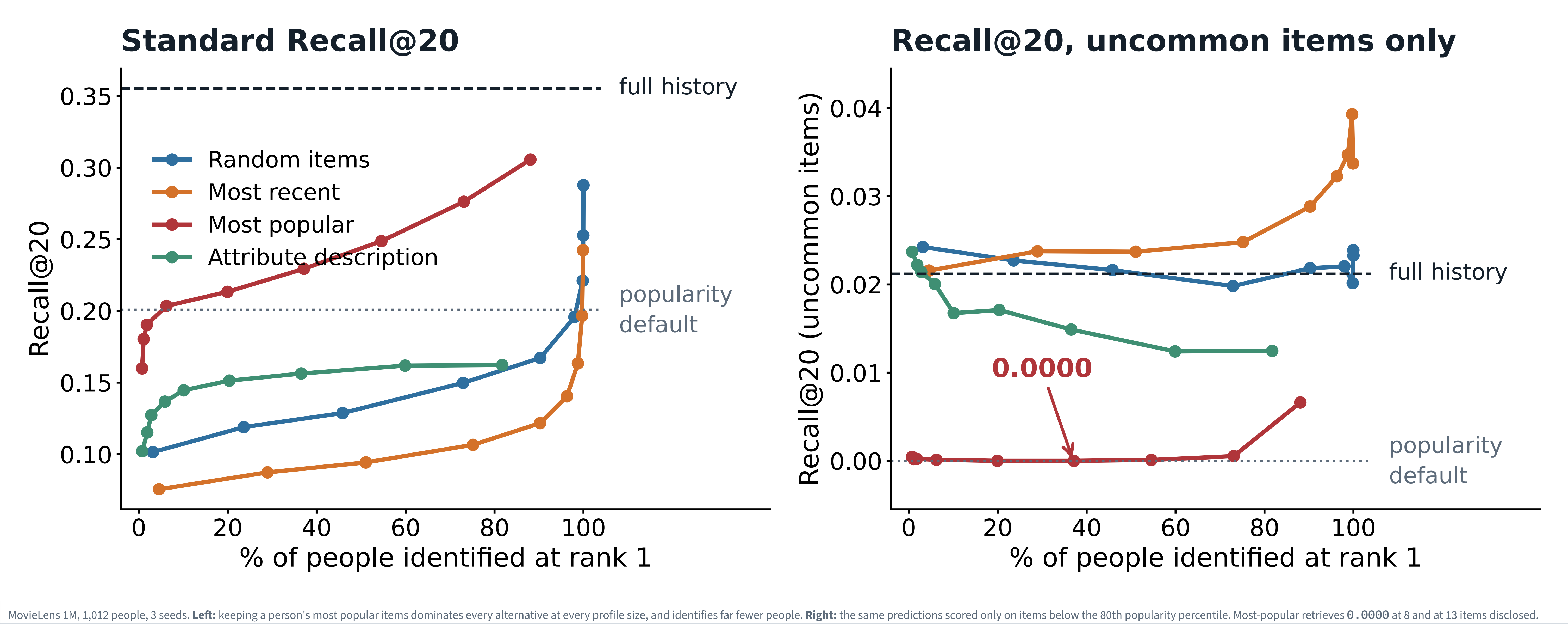
Measuring the point where more context stops helping

Moses Nderemani · Pervasive Interaction Research Group, Tampere University

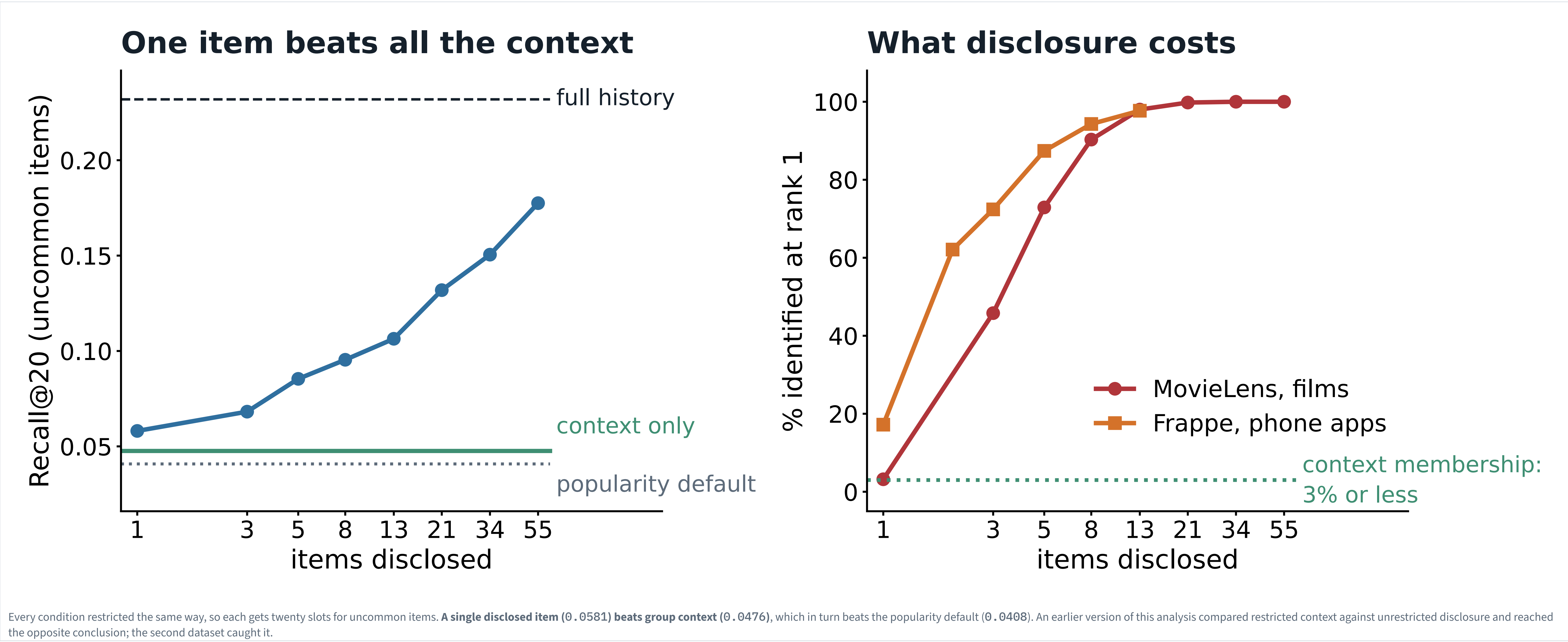
SDLCAI 2026 · Marsio, Aalto University · 13 October

The data-minimisation strategy that scores best on the standard metric retrieves **zero** uncommon items. It wins by abandoning the person.

THE QUESTION	WHAT WAS KNOWN	WHAT IS ADDED HERE
Personalised systems accumulate data, most of it observed rather than volunteered. GDPR Article 5(1)(c) requires that it be limited to what is necessary. How little is that?	Biega et al. (SIGIR 2020) formalised this as performance-based data minimisation, compared seven ways of choosing what to keep, and found the accuracy cost modest but unevenly spread across people.	Quality and re-identification measured together rather than separately, on two datasets, with a popularity control that the earlier framing had no reason to run.



0.0000 Uncommon items found by the best-scoring strategy. Told only about someone's popular items, the model returns something close to a bestseller list. Scored against targets that are themselves mostly popular, that looks like success.	-0.284 Rank correlation between the standard metric and the uncommon-item metric across 45 strategy and size combinations ($p = 0.058$, short of significance). The direct comparison carries the claim: 0.0000 against 0.0221 at 13 items.	90.3% Of people identified at rank 1 from 8 disclosed films, matched against all 1,012 profiles, without timestamps. On Frappe, 2 phone apps identify 62.1%. Informative and identifying are the same property.
--	---	---



Which cue you pick matters more than whether you have one		WHAT TO BUILD
weather	65%	• Use context to replace the default, not the profile. It is a strictly better version of what a system shows when it knows nothing, at 3% identification cost or less.
time of day	35%	• Reserve capacity for the uncommon. The largest single effect here was refusing to fill a list with popular items, which people already have access to.
weekday	15%	• Choose the cue empirically. The spread between context variables (~2% to +65%) is wider than the gap between having context and not having it.
weekend?	12%	• Treat a missing cue as missing. A signal that is usually absent is worse than none, because the system falls back silently while appearing to adapt.
home / work	-2%	
% gain over the popularity default		

Frapple, 957 users, 4,082 apps; cohort 87 users and 235 user-context cells. Home or work is 79% unknown in the data, and gives nothing.